



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

**Sistema de Informação Gerencial para o Programa de
Iniciação Científica (ProIC) da Universidade de
Brasília (UnB)**

Roberto Sobrinho da Silva
Mario Matos Camargo Filho

Monografia apresentada como requisito parcial
para conclusão do Curso de Computação — Licenciatura

Orientadora
Prof.^a Dr.^a Maristela Terto de Holanda

Brasília
2013

Universidade de Brasília — UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Curso de Computação — Licenciatura

Coordenador: Prof. Dr. Flávio de Barros Vidal

Banca examinadora composta por:

Prof.^a Dr.^a Maristela Terto de Holanda (Orientadora) — CIC/UnB
Prof.^a Dr.^a Aletéia Patrícia Favacho de Araújo — CIC/UnB
Prof.^a Dr. Mário César Ferreira — IP/UnB

CIP — Catalogação Internacional na Publicação

Silva, Roberto Sobrinho da.

Sistema de Informação Gerencial para o Programa de Iniciação Científica (ProIC) da Universidade de Brasília (UnB) / Roberto Sobrinho da Silva, Mario Matos Camargo Filho. Brasília : UnB, 2013.

201 p. : il. ; 29,5 cm.

Monografia (Graduação) — Universidade de Brasília, Brasília, 2013.

1. Iniciação Científica; Gestão da Informação; Inteligência de Negócios; Armazém de Dados; Modelagem Dimensional e Extração,
2. Transformação e Carga

CDU 004.4

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro — Asa Norte
CEP 70910-900
Brasília-DF — Brasil

Dedicatória

Dedicamos este trabalho à Universidade de Brasília (UnB), especificamente ao Decanato de Pesquisa e Pós Graduação (DPP) para o qual este trabalho foi desenvolvido, com a expectativa de contribuir com gestão da Iniciação Científica da Universidade.

Agradecimentos

Agradecemos aos nossos familiares, em especial aos nossos pais: Edna Rosa Sobrinho da Silva e Rodemário Moreira da Silva, pais de Roberto, Marina Conceição de Rezende Camargo e Mario Matos Camargo, pais de Mario, pois são a principal razão do nosso sucesso. A nossa orientadora Maristela Terto de Holanda pela disponibilidade, suporte, profissionalismo, generosidade e dedicação. Ao Dr. Mario César Ferreira pela idealização do projeto de qual este trabalho faz parte e pela oportunidade. Aos amigos que nos apoiaram ao longo desta etapa.

Abstract

The Implementation of Management Information System for the Scientific Initiation Program at University of Brasilia (UnB) demands a different approach to the development of a database system that runs on a Web server. The Business Intelligence is a fundamental concept to the development of a Data Warehouse because it determines the necessity of finding the best possible solution for a certain type of problem. Through multi-dimensional modeling is implemented a Data Warehouse that guarantees queries with good quality. Through a Online Process Analytical Tool, can be built multiple objects (reports, graphs, tables) that is important for a robust analysis of a data. The Pentaho Platform is formed by the necessary tools for the construction of a data warehouse system, supporting the implementation of all the steps required to build the multidimensional model, implementation extraction, transformation and loading process, creation of analytical cubes and generation of Web Ad-Hoc Query Reports.

Keywords: *Scientific Initiation; Management Information; Business Intelligence; Data Warehouse; Dimensional Modeling and Extract, Transform and Load*

Resumo

A implementação do Sistema de Informação Gerencial para o Programa de Iniciação Científica (ProIC) da Universidade de Brasília (UnB) demanda uma abordagem diferente para o desenvolvimento de um sistema de Banco de Dados que rode em um servidor *Web*. A Inteligência de Negócios é um conceito fundamental para o desenvolvimento de um Armazém de Dados, pois determina que deve-se buscar a melhor solução possível para um certo tipo de problema. Por meio da modelagem multidimensional, pode-se implementar um Armazém de Dados que garanta buscas com boa qualidade. Através de uma Ferramenta que interpreta Processos Analíticos *Online*, podem ser construídos vários objetos (relatórios, gráficos, tabelas) importantes para uma análise robusta destes dados. A Plataforma Pentaho é formada pelas ferramentas necessárias para a construção de um sistema de Armazém de Dados, dando suporte a implementação de todas as etapas necessárias como a construção do modelo multidimensional, o processo de extração, transformação e carga, a criação de cubos analíticos e geração de relatórios Ad-Hoc.

Palavras-chave: Iniciação Científica; Gestão da Informação; Inteligência de Negócios; Armazém de Dados; Modelagem Dimensional e Extração, Transformação e Carga

Sumário

1	Introdução	1
1.1	Justificativa	2
1.2	Objetivos	2
1.3	Estrutura do Trabalho	2
2	A Iniciação Científica	4
2.1	Programa Institucional de Bolsas de Iniciação Científica (PIBIC)	4
2.2	Programa de Iniciação Científica da UnB (ProIC)	5
2.3	Dados do Sistema de Bolsas da Iniciação Científica da UnB	7
3	Revisão Teórica	10
3.1	<i>Business Intelligence</i> (BI)	10
3.2	<i>Data Warehouse</i> (DW)	11
3.2.1	Sistemas Transacionais e Sistemas Analíticos	13
3.3	Modelo Multidimensional	14
3.3.1	Granularidade	16
3.3.2	Dimensões	16
3.3.3	Métricas	16
3.4	<i>Extract, Transform and Load</i> (ETL)	17
3.5	<i>Online Analytical Processing</i> (OLAP)	18
4	Plataforma Pentaho	21
4.1	Visão Geral do Pentaho	21
4.2	Integração de Dados no Pentaho	23
4.3	Servidor BI	24
4.4	Mondrian, MDX e <i>Pentaho Schema Workbench</i>	29
4.5	<i>Pentaho Metadata Editor</i> (PME)	29
5	Implementação	31
5.1	Análise (Diagnóstico) do Processo	31
5.2	Modelagem Multidimensional	31
5.3	Processo de ETL	32
5.4	Criação do DW	36
5.5	Cubo OLAP MDX	37
5.6	Metadados e Relatórios WAQR	37

6	Resultados	40
6.1	O Sistema de Informação Gerencial	40
6.2	Qualidade dos Dados	43
6.2.1	Análise dos Registros dos Projetos e Planos de Trabalho	44
6.3	Relatórios Gerenciais	45
6.3.1	Perfil dos Alunos	46
6.3.2	Perfil dos Orientadores	47
6.3.3	Bolsistas no Período	49
6.3.4	Número de Voluntários no Período	50
7	Conclusão	53
	Referências	55
A	Relatório WAQR Título do Orientador por Área e Ano	58
B	Relatório WAQR da Média do IRA pelo Ano	71
C	Mapa Lógico de Dados	76

Lista de Figuras

2.1	Organograma do Decanato de Pesquisa e Pós Graduação.	6
2.2	Sistema de Bolsas da Iniciação Científica (SiBol).	8
2.3	Sistema de Bolsas da Iniciação Científica Online (SiBolWeb).	9
3.1	DW como Item Central na Solução de um BI [31].	12
3.2	Modelo Multidimensional (Ligação entre o Fato e as Dimensões) [2].	15
3.3	Cubo OLAP Tridimensional [2].	20
4.1	O <i>Workflow</i> da Plataforma Pentaho.	22
4.2	<i>Steps Big Data e Input</i>	25
4.3	<i>Steps Data Warehouse, Validation, Statistics, Palo, Job, Mapping, Bulk Loading, Inline, Experimental e Deprecated</i>	26
4.4	<i>Steps Utility, Flow, Scripting e Transform</i>	27
4.5	<i>Steps Scripting, Lookup e Joins</i>	28
4.6	Processo de Criação de Metamodelos.	30
5.1	PDI - Extração.	33
5.2	Transformação da Dimensão Projeto.	34
5.3	Transformação da Dimensão Aluno do Banco de Dados Access.	35
5.4	Transformação do Fato.	35
5.5	<i>Dimension Lookup/Update</i>	36
5.6	Cubo Vinculando a Tabela Fato_Plano.	38
5.7	Fonte de Relatórios WAQR.	39
6.1	Tela de <i>Login</i>	41
6.2	Área de Trabalho.	41
6.3	Tela de Criação de WAQR.	42
6.4	Tela de Relatório WAQR.	42
6.5	Tela de Navegação OLAP.	43
6.6	Gênero por Período.	46
6.7	Quantidade de Discentes em Cada Campus por Período.	49
6.8	Tipo de Bolsa por Período.	51

Lista de Tabelas

3.1	Diferenças entre os Sistemas OLTP e OLAP.	13
3.2	<i>Slice</i> por Produto [2].	19
3.3	<i>Dice</i> entre Revenda e Tempo [2].	19
3.4	<i>Drill-Across</i> entre Revenda e Produto [2].	19
4.1	Vantagens da Plataforma Pentaho [33].	22
5.1	Planilha Mapa Lógico dos Dados.	32
6.1	Coerência das Chaves nos Registros dos Orientadores.	44
6.2	Matrículas e CPFs de Orientadores.	44
6.3	Quantidade de Planos de Trabalho e Quantidade de Alunos.	45
6.4	Gênero por Período.	47
6.5	Quantidade de Discentes em Cada Campus por Período.	50
6.6	Tipo de Bolsa por Período.	52
C.1	Gestão	76
C.2	Identificação do DW	77
C.3	Fontes (Tabela e Tipo)	80
C.4	Fontes (Tamanho e Precisão)	86
C.5	Transformação	93

Capítulo 1

Introdução

O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) é uma agência vinculada ao Ministério da Ciência, Tecnologia e Inovação (MCTI) do Governo Federal que tem como objetivo principal estimular a pesquisa, o desenvolvimento de tecnologias e a formação de pesquisadores. Por meio do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) realiza parcerias com as principais organizações de ensino superior e pesquisa do país (públicas e privadas) ofertando bolsas de Iniciação Científica (IC) a alunos de graduação vinculados a esses institutos. Na Universidade de Brasília (UnB) o órgão responsável pela organização e seleção dos projetos de IC é o Decanato de Pesquisa e Pós Graduação (DPP), vinculado a reitoria da universidade.

O Centro de Informática (CPD), órgão que atua na área de Tecnologia da Informação (TI) da UnB, foi responsável pelo desenvolvimento dos sistemas (inicialmente o Sibol e posteriormente o SibolWeB) que gerenciam as informações vinculadas a atividade de fomento da IC na UnB. A perspectiva de aumento do número de projetos de IC ao longo do tempo exige uma nova solução para gerenciamento dessas informações. Nesta perspectiva, a inteligência de negócios é de grande valia para se obter informações consolidadas para apoio de decisão gerencial eficiente, com rapidez e disponibilidade. Uma boa gestão de processos é fundamental porque propicia ao gestor uma maior consciência sobre os eventos envolvidos e a influencia deles na dinâmica do processo. Uma importante ferramenta para os gestores é o conhecimento e o domínio da dimensão desses eventos, pois a probabilidade de se tomar decisões precisas e coerentes aumenta com a posse de informações significativas.

Sistemas Transacionais são úteis para garantir a inserção dos dados, mantendo o estado do banco de dados consistente, porém são pouco eficientes para lidar com um grande volume de dados, dificultando a criação de uma série histórica de informações. Uma análise objetiva, detalhada e rica de um universo que envolve uma grande base de dados só é possível caso seja viável extrair informações por meio do confrontamento de uma quantidade robusta de registros. Os sistemas analíticos, também conhecidos como *Online Analytical Processing* (OLAP), são mais eficientes na tomada de decisões e no diagnóstico global de grandes volumes de dados. [33]

Com a criação e a carga de um armazém de dados (*Data Warehouse*), é possível criar um sistema de banco de dados capaz de mapear o histórico da IC na UnB, e gerar relatórios que auxiliam o DPP no planejamento da política de incentivo à pesquisa científica. Outra etapa importante é o processo de extração, tratamento e carga de dados, e a transcrição

dos sistemas transacionais para sistemas analíticos com o auxílio da técnica de modelagem dimensional.

1.1 Justificativa

O DPP é o responsável pela gerência da Iniciação Científica na UnB e para isso atualmente existe um sistema, o SibolWeb, que auxilia nas inscrições e controle dos projetos, cadastro dos orientadores e alunos. O SibolWeb começou a operar em 2004 com persistência em *MySQL*, e antes deste sistema havia bases de dados em Access que eram feitas em arquivos distintos, anualmente de 2000 até 2003, oriundas do Sibol.

Nesta monografia propõe-se a criação de uma solução de informações gerenciais da iniciação científica da UnB para registro e tratamento de dados, e geração de relatórios técnicos com a definição sistemática de tratamento estatístico dos indicadores presentes no banco de dados para a análise e formulação de políticas de fomento da iniciação científica na UnB.

O uso dos conceitos *Business Intelligence* (BI) e *Data Warehouse* (DW) conjuntamente, possibilita a extração de informações que ajudam a melhorar a gestão da iniciação científica. As boas práticas de gerenciamento encontradas em BI e a grande capacidade de análise de dados de um DW são essenciais para o desenvolvimento de um sistema que auxilie o DPP na gerência dos projetos de IC. Uma solução em inteligência de negócios aplicado a gerência dos projetos de IC da UnB, possibilita que o DPP tenha acesso a uma análise histórica evolutiva dos projetos de IC desenvolvidos na universidade.

Atualmente, a Diretoria de Fomento à Iniciação Científica (DIRIC), responsável pela gestão do PIBIC na UnB, não possui nenhuma ferramenta que auxilia o gestor na tomada de decisão. Acredita-se, então, que essa ferramenta vai trazer melhoria para a gestão das bolsas de IC coordenadas pela UnB. Também é importante destacar que com uma quantidade abrangente de indicadores quantitativos disponíveis, resultado do crescimento da IC, uma ferramenta como essa ajudará ao DPP traçar o perfil da IC na instituição nos últimos anos.

1.2 Objetivos

Implementar um sistema de informação gerencial para o Decanato de Pós Graduação e Pesquisa da UnB relacionados com os dados da iniciação científica ao longo dos últimos doze anos. Dentre os objetivos específicos tem-se:

- a modelagem do DW;
- definição do processo de carga dos dados do DW;
- criação de cubos para o processo de análise;
- disponibilizar a solução em um servidor BI.

1.3 Estrutura do Trabalho

O desenvolvimento deste trabalho será apresentado pelos seguintes capítulos:

- Capítulo 2 , tem como objetivo apresentar a iniciação científica, assim como os seus dados históricos.
- Capítulo 3 , apresenta o referencial teórico fundamental para o trabalho. Dentre os termos tem-se: *Business Intelligence* (BI), *Data Warehouse*, *Extract, Transform and Load* (ETL) e o Modelo Multidimensional.
- Capítulo 4 , descreve as funcionalidades de implementação da plataforma Pentaho e descreve cada aplicativo utilizado neste trabalho.
- Capítulo 5 , mostra a implementação da solução de BI, desde a análise dos bancos de dados até a produção de relatórios. Todas as etapas da implementação do sistema são descritas na seguinte ordem: análise do processo e da base de dados, modelagem dos dados, o processo ETL, implmentação dos cubos OLAP e consultas MDX.
- Capítulo 6 , apresenta os relatórios e os resultados de todo o projeto.
- Capítulo 7 , traz a conclusão deste trabalho.

Capítulo 2

A Iniciação Científica

Este capítulo apresenta o Programa Institucional de Bolsas de Iniciação Científica (PIBIC), como se desenvolve a iniciação científica na Universidade de Brasília. Em seguida, alguns dados do sistema que gerencia essas bolsas na UnB são apresentados.

2.1 Programa Institucional de Bolsas de Iniciação Científica (PIBIC)

O Programa Institucional de Bolsas de iniciação científica é vinculada ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e tem como principal objetivo consolidar políticas de iniciação científica a serem desenvolvidas pelas Instituições de Ensino Superior e Pesquisa, através da concessão de bolsas de IC, a estudantes de graduação, introduzindo-os na pesquisa científica [1].

As bolsas de IC são oferecidas, pelo CNPq, diretamente às instituições, que ficam responsáveis pela seleção dos projetos dos pesquisadores orientadores que tem interesse em participar do PIBIC. No caso dos estudantes, as bolsas são ofertadas por meio da indicação dos orientadores. As metas do Programa de IC são:

- despertar vocação científica e incentivar os estudantes de graduação a se tornarem pesquisadores;
- acelerar o processo de formação de titulação de novos mestres e doutores;
- contribuir para a formação científica de recursos humanos em diversas áreas;
- criar uma maior articulação entre a graduação e pós-graduação;
- formar recursos humanos para área de pesquisa;
- reduzir o tempo de permanência dos alunos na pós-graduação;
- estimular os pesquisadores produtivos a direcionar os alunos de graduação em atividades científica, tecnológica e artístico-cultural;
- proporcionar ao bolsista, orientado por pesquisador qualificado, a aprendizagem de técnicas e métodos de pesquisa;

- desenvolver o pensamento científico e criativo nos bolsistas, a partir de condições criadas pelo confronto direto com os desafios de pesquisa;
- ampliar o acesso do estudante à cultura científica.

Uma bolsa tem validade de doze meses, a partir do primeiro mês de vigência do processo institucional, com início no dia primeiro de agosto. A bolsa de IC para o estudante de graduação é correspondente ao valor atual fixado pela tabela de valores de bolsas no País. Os requisitos e as condições necessários para participar do PIBIC são:

- para as instituições:
 - requisitos: instituições públicas, comunitárias ou privadas, com ou sem curso de graduação, que efetivamente desenvolvem pesquisa e possuem infraestrutura adequada.
 - condições de participação: a chamada pública de propostas para o processo de inscrição ocorre no primeiro semestre de cada ano, em geral, entre os meses de março, abril e é publicada no item Editais da página do CNPq.
- para os pesquisadores e orientadores:
 - requisitos: possuir vínculo com a instituição de Ensino e Pesquisa que participa do PIBIC e desenvolver pesquisa científica e, preferencialmente, possuir vínculo de Produtividade em Pesquisa com o CNPq.
 - condições de participação: participar em processo de seleção realizado na instituição de vínculo. Os processos de seleção nas instituições ocorrem, em geral, no primeiro semestre de cada ano.
- para os estudantes:
 - requisitos: cursar graduação, e estar disponível para as atividades acadêmicas e de pesquisa.
 - condições de participação: buscar, na área de interesse, um pesquisador disposto a integrar o bolsista em sua pesquisa e a orientá-lo.

2.2 Programa de Iniciação Científica da UnB (ProIC)

O Decanato de Pesquisa e Pós Graduação (DPP), órgão da Reitoria, é responsável pela formulação, coordenação e cumprimento do Plano de Desenvolvimento Institucional da Universidade de Brasília, no que tange à sua política de pós graduação, pesquisa e desenvolvimento em todas as áreas de conhecimento. O DPP possui a missão de promover, coordenar, supervisionar e apoiar o ensino de pós-graduação e a pesquisa, com o objetivo de formar recursos humanos de alto nível, visando o crescimento, a disseminação e a internacionalização da pós-graduação, da pesquisa e da produção de conhecimento necessária ao desenvolvimento científico, tecnológico, artístico e cultural do país [3].

A história da IC na UnB não é recente. Ela se estrutura oficialmente a partir de 1994, após diversas iniciativas (seminários) que buscavam, fundamentalmente, aprimorar o ensino de graduação e estreitar os laços da Pós Graduação (pesquisa) com os alunos

da graduação (ensino). Seis anos antes, em 1988, o CNPq já havia criado o PIBIC. Isto contribuiu fortemente para impulsionar a IC em todo o Brasil [6].

Ao longo de 18 anos, a IC na UnB se consolidou como política institucional de fomento à pesquisa na graduação, integrando professores-pesquisadores e alunos das três grandes áreas do conhecimento: Artes e Humanidades, Ciências da Vida e Ciências Exatas e Tecnológicas.

A consolidação da IC na UnB se expressa por meio de diversos indicadores, os quais se destacam [3]:

- realização de dezoito Congressos de IC (CICs);
- apresentação de cerca de 17.000 trabalhos nos CICs;
- participação em eventos científicos, principalmente, nas reuniões anuais da SBPC;
- crescimento dos investimentos em bolsas de iniciação científica.

Essa monografia visa dar suporte ao projeto de Pesquisa do professor Dr. Mário César Ferreira, diretor de fomento da iniciação científica na UnB no período de 2009 a 2012, cargo vinculado à Diretoria de Fomento à iniciação científica (DIRIC) do DPP [6]. A Figura 2.1 [34] mostra o organograma dos órgãos vinculados ao DPP da UnB.

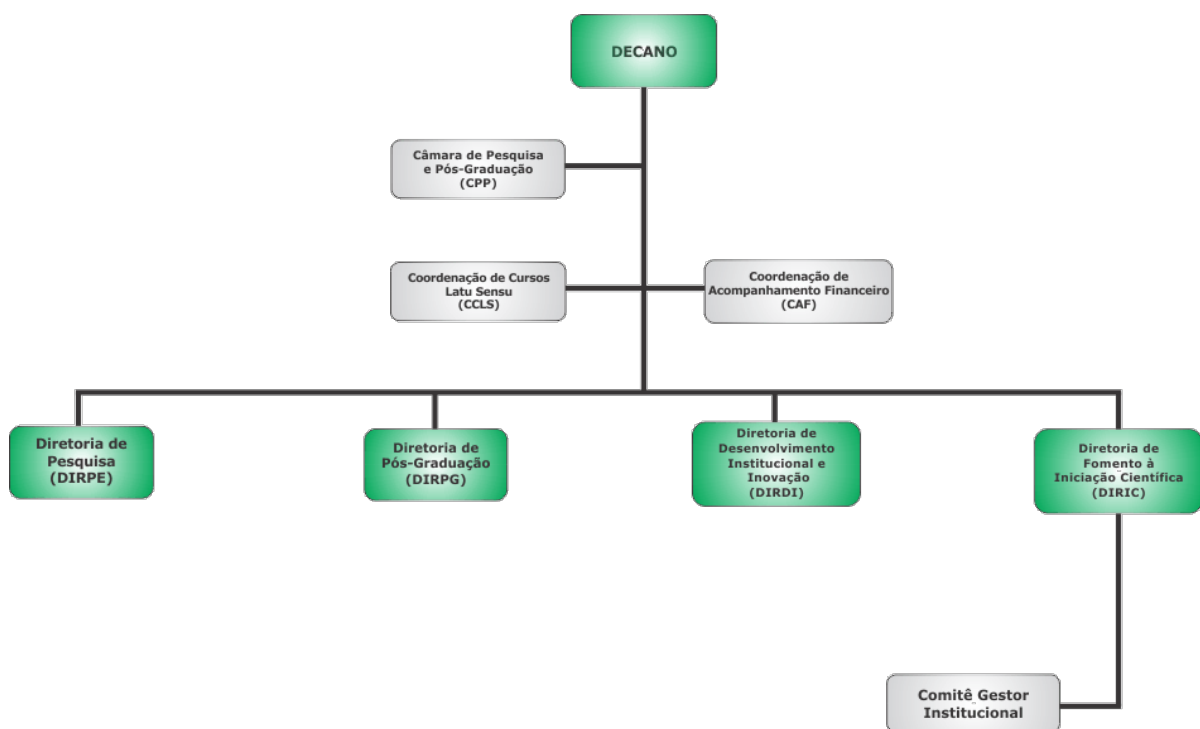


Figura 2.1: Organograma do Decanato de Pesquisa e Pós Graduação.

O fomento à IC na UnB é, sem dúvidas, uma iniciativa bem sucedida. Todavia, sua evolução sustentável requer transpor diversos limites existentes, como ferramentas computacionais para auxiliar na definição de políticas internas.

As modalidades essenciais para o estudo de IC na UnB são três [3]:

- social: o desenvolvimento regional sustentável a partir do alinhamento do perfil da produção científica em IC da UnB, e as especificidades regionais do Distrito Federal e região do entorno.
- acadêmica: o incentivo à produção científica e bibliográfica da UnB por meio do conhecimento rigoroso do perfil e evolução da IC na história da UnB.
- institucional: o crescimento de IC na UnB por meio de um levantamento abrangente de indicadores quantitativos que tracem o perfil da IC na instituição.

A pertinência institucional da investigação se prende ao resgate da história documental, institucional e na definição de indicadores institucionais quantitativos nas áreas de conhecimento, no perfil de docente e discente.

Nesse cenário, propõe-se a criação de uma solução de informações gerenciais para registro, tratamento de dados e geração de relatórios técnicos com a definição sistemática de tratamento estatístico dos indicadores presentes no banco de dados para a análise e formulação de políticas de fomento da iniciação científica na UnB.

O foco de investigação para a iniciação científica no contexto institucional da Universidade de Brasília tem como base dois eixos interdependentes [3]:

- eixo histórico: buscar do conhecimento, com rigor científico, dos principais fatos, personagens e espaços institucionais implicados na evolução temporal da IC na UnB.
- eixo indicadores: definir variáveis quantitativas, conceber o banco de dados de IC e estabelecer o tratamento estatístico padronizado para a geração de relatórios técnicos e sintéticos.

As características da IC na Universidade de Brasília requerem uma abordagem interdisciplinar de investigação que permita que a DIRIC conheça a história e o perfil dos indicadores quantitativos, fruto dos levantamentos estatísticos realizados até este momento, que caracterizam a IC na UnB. O conhecimento mais apurado desses dados permite que a reitoria da Universidade de Brasília [3]:

- identifique os principais fatos, personagens e espaços institucionais presentes na evolução histórica da IC na UnB;
- defina as melhores políticas de pesquisa científica para UnB incorporando de maneira definitiva, a solução de informações gerenciais, sobre os bancos de dados existentes nos sistemas de controle do processo da IC;
- estabeleça padrão de tratamento estatístico e geração de relatórios técnicos;
- realize o mapeamento constante do perfil histórico e evolutivo da IC na UnB;
- mantenha contínua a prática de levantamento de dados da IC na UnB.

2.3 Dados do Sistema de Bolsas da Iniciação Científica da UnB

No ano de 2000, entrou em funcionamento o primeiro sistema de controle de dados relativo aos projetos de IC da UnB, Sistema de Bolsas da iniciação científica (SiBol). O

sistema funcionava a partir de modelos relacionais de bancos de dados (formato mdb), criados com a ferramenta Microsoft Access [19], parte da plataforma Microsoft Office. O primeiro banco de dados foi gerado a partir dos dados da IC de 2000. Nos anos seguintes em que o sistema vigorou, entre 2001 a 2003, foram criados os outros modelos de bancos de dados Access, atualizados a partir do acréscimo de novas tabelas, funcionalidades e novos recursos. No período de 2000 até 2003 eram desenvolvidos, para cada ano, um Banco de dados próprio que englobava somente as informações das bolsas de IC referentes ao edital do ano. A Figura 2.2 mostra o *layout* do sistema SiBol.

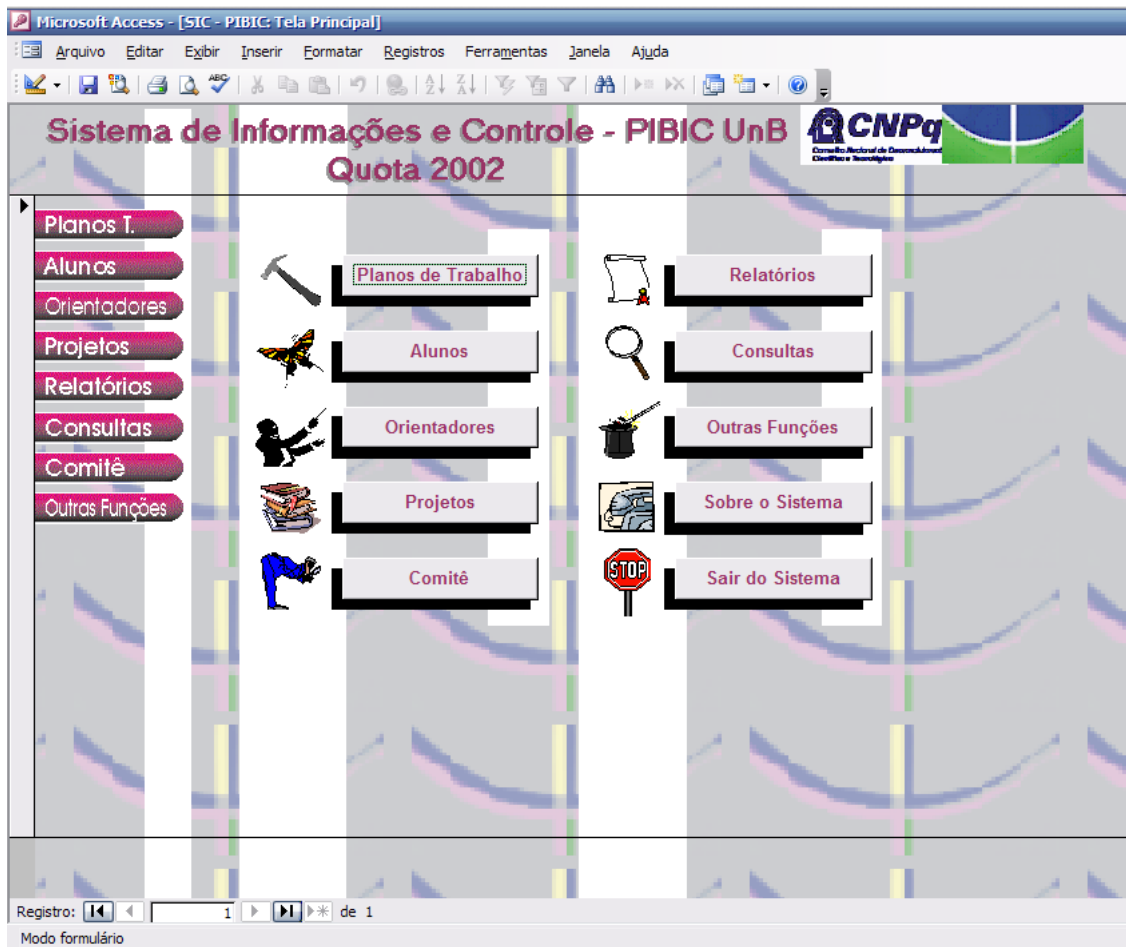


Figura 2.2: Sistema de Bolsas da Iniciação Científica (SiBol).

A partir do ano de 2004 foi desenvolvida uma nova versão para a plataforma de Sistema de Bolsas (SiBolWeb), que ficou disponível na *web* por meio do endereço <https://www.sibolweb.unb.br/SibolWeb/Fronteira/formlogin.aspx>, dentro do domínio unb.br, e continua sendo utilizada nos dias atuais. A partir dessa versão, o cadastro de projetos de IC é feita por meio de cadastro *online*, carregados em uma base de dados única, diferente do procedimento adotado pela versão anterior do sistema de bolsas. A plataforma atual foi desenvolvida com uso da linguagem *Hypertext Preprocessor* (PHP), em conjunto com um Sistema Gerenciador de Banco de Dados (SGBD) *MySQL*. Essa evolução do sistema tornou o cadastro de bolsas de IC mais dinâmico, eficiente e prático, e permitiu a melhora da gestão das bolsas de IC na UnB. A Figura 2.3 mostra o *layout* do sistema SiBolWeb.

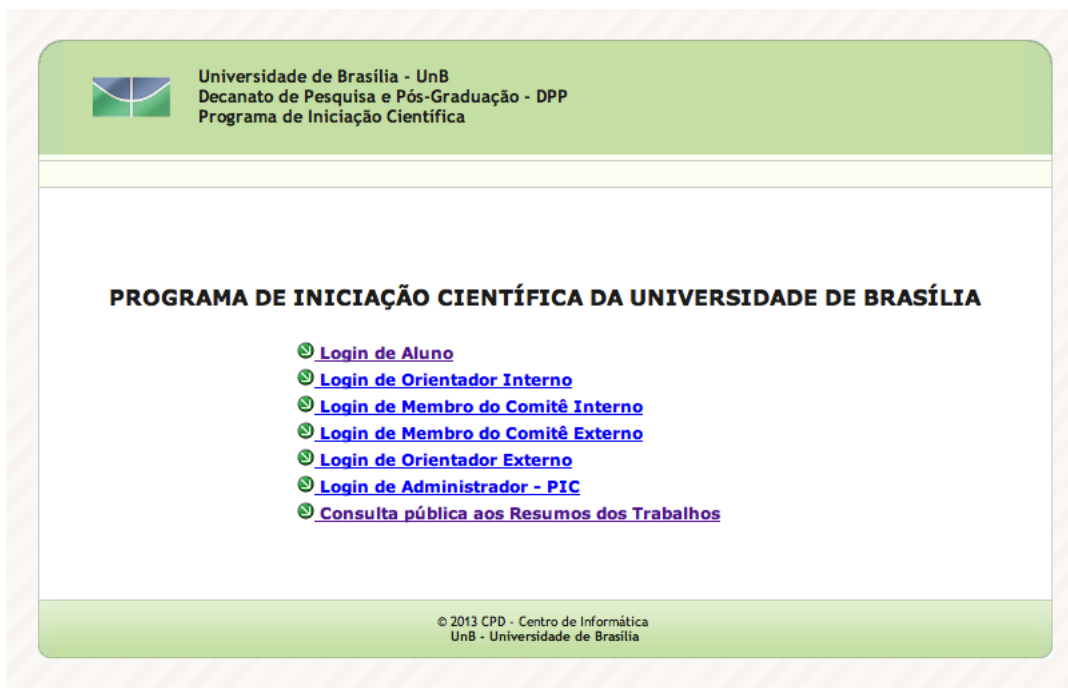


Figura 2.3: Sistema de Bolsas da Iniciação Científica Online (SiBolWeb).

Na gestão do DPP, no período de 2009 a 2012, a decana Prof^a. Dr^a. Denise Bomtempo Birche de Carvalho, em conjunto com a DIRIC, cujo titular era o Prof. Dr. Mário César Ferreira, adotaram como um de seus objetivos, implantar soluções que me-lhorassem a qualidade da gestão da IC na UnB. O DPP, em conjunto com Prof^a. Dr^a. Célia Ghedini Ralha, titular da Diretoria de Desenvolvimento Institucional e Inovação (DIRDI), que faz parte do quadro de professores do Departamento de Ciência da Computação (CIC), já havia firmado uma parceria com o CIC com a finalidade de avaliar e compreender os processos de trabalho do DPP. Para propor melhorias e automatizações foram definidos dez grupos de alunos de graduação, na disciplina de Sistemas de Informação, do período 2010/1, ministrada pela Prof^a. Célia Ghedini Ralha. de IC.

A partir das necessidades atuais do DPP quanto à evolução do sistema SiBolWeb, surgiu a oportunidade para o desenvolvimento deste projeto final de graduação. Uma das necessidades do DPP é que a nova versão do sistema de bolsas permita a integração de todos os dados da IC nos anos de 2000 a 2012 em uma base de dados única, e também suporte a adição de novos registros referentes aos anos posteriores, sem que o desempenho do sistema seja afetado pela tendência de aumento de volume de dados. Além disso é muito importante que sejam incluídas novas funcionalidades que permitam, aos usuários da nova versão do sistema, a criação de relatórios, gráficos e tabelas para um melhor diagnóstico da situação da IC na UnB. O conceito de inteligência de negócios e a construção de um armazém de dados são as soluções básicas que foram aplicadas no desenvolvimento deste novo sistema de banco de dados.

Capítulo 3

Revisão Teórica

Este capítulo apresenta os conceitos fundamentais para o desenvolvimento deste trabalho, sendo dividido nas seguintes seções: a Seção 3.1 apresenta o conceito de *Business Intelligence* (Inteligência de Negócios), a Seção 3.2 apresenta o conceito de *Data Warehouse* (Armazém de Dados), a Seção 3.3 apresenta as características do modelo multi-dimensional, a Seção 3.4 apresenta o conceito de *Extract, Transform and Load* que é o processo de extração, transformação e carga do banco de dados, e a Seção 3.5 apresenta o conceito de *Online Analytical Process* (OLAP) utilizado na geração de relatórios.

3.1 *Business Intelligence* (BI)

Para uma boa gestão de processos é fundamental que o gestor tenha consciência sobre os eventos envolvidos e os que influenciam na dinâmica do negócio. Uma importante ferramenta para os gestores é o conhecimento e o dimensionamento desses eventos, pois tendo informações significativas as decisões serão mais precisas e coerentes.

Segundo Swain Scheps [28], *Business Intelligence* (BI) é qualquer atividade, ferramenta ou processo usado para obter a melhor informação para apoiar o processo de tomada de decisões.

A inteligência empresarial é outra denominação relacionada a BI. O conceito descreve as habilidades das corporações para acessar dados e explorar as informações, analisando-as e desenvolvendo percepções e entendimentos a seu respeito. Isto permite incrementar e tornar mais pautada em informações as tomadas de decisão.

A inteligência de negócios pode ser um software que possibilita aos usuários obter informações corporativas facilmente [5]. Esse tipo de *software* deve permitir que o usuário consiga utilizar o sistema sem que haja um conhecimento técnico sobre banco de dados, linguagem SQL ou mesmo como os dados estão organizados dentro do banco. Uma arquitetura padrão de soluções de BI possui três componentes básicos: o repositório de dados não volátil (DW), o processo de extração de dados (ETL) e a área de apresentação.

O termo BI tem várias definições, pois se leva em consideração o campo de discussão e o seu contexto, e naturalmente eles se permeiam. Na computação, BI é uma tecnologia com abordagem em várias áreas como banco de dados e engenharia de software. Na administração, uma ferramenta de gestão.

Em todas as empresas, o processo de criação de negócios é composto por uma série de eventos que caracterizam suas atividades principais. A natureza e frequência destes even-

tos variam conforme o tipo de negócio em que uma empresa está envolvida, ou seja, um produto é manufaturado, uma conta é creditada enquanto outra é debitada, um assento é reservado, um pedido é incluído, etc. O controle e o processamento corretos destes eventos são críticos para uma empresa, sendo que estas atividades contribuem para seu sucesso ou fracasso. Para isso, muitas organizações possuem um conjunto de sistemas conhecidos como sistemas transacionais ou sistemas *Online Transaction Processing* (OLTP) que capturam os eventos de forma individual e todos os detalhes associados a eles. Cada um destes sistemas está encarregado de um tipo diferente de atividades, e trata das transações de negócios segundo um conjunto de regras que garanta sua consistência e o armazenamento de todos os detalhes associados. Para um sistema OLTP, os princípios de desenho como normalização e consistência de transações são extremamente críticos [12].

3.2 *Data Warehouse* (DW)

O *Data Warehouse* (DW) surgiu com a crescente demanda de sistemas de apoio à decisão (SADs), pois os sistemas transacionais são focados na operacionalização do negócio e a armazenam em seus Sistemas de Gerenciamento de Banco de Dados (SGBDs) referentes à operacionalização da rotina do negócio. Segundo Ramez Elmasri e Shamkant B. Navathe [27], há grande necessidade de proporcionar aos tomadores de decisão, da gerência média para cima, informações no nível correto de detalhe para dar suporte à tomada de decisão.

Um DW existe para responder as questões que as pessoas tem sobre os negócios. Esta função contrasta fortemente com o propósito dos sistemas transacionais que as empresas utilizam e requer que o desenho ou o modelo de dados do DW siga princípios completamente diferentes. As técnicas de modelagem dimensional, se aplicadas corretamente, garantem que o projeto do DW reflita a forma de pensar dos gerentes de negócio e possam ser utilizadas para atendê-los [12].

Segundo Visek R. Gupta [10], com o surgimento dos DWs foi necessário criar novos métodos de estruturação de dados, no armazenamento e na consulta da informação. As empresas produzem e armazenam um volume elevado de dados, sendo normal que estes dados estejam dispersos por vários servidores, que podem, inclusive, estar dispersos por várias localizações geográficas e ter sido desenvolvidos em plataformas e linguagens diferentes.

Os DWs possuem a característica distinta de que são direcionados, principalmente, para aplicações de apoio às decisões. Eles são otimizados para a recuperação de dados, não para o processamento rotineiro de transações [27].

Um armazém de dados não é uma aplicação que se compra e instala nos computadores da empresa. Na realidade, a sua implementação exige a integração de vários produtos e processos. Numa perspectiva minimalista, um DW não é mais que uma base de dados especializada, que integra e gera o recolhimento de informações a partir de sistemas transacionais internos e fontes de dados externas. Neste contexto, um DW é construído para permitir, uma visão integrada e completa de toda a organização, acesso aos dados históricos da organização, tendo uma fonte de dados verossímeis dentro da organização e facilitando os processos de tomada de decisão, sem sobrecarregar os sistemas operacionais [26].

Os processos e negócios são dinâmicos e mudam ao longo do tempo incluindo o entendimento sobre os dados e informações armazenadas nos sistemas operacionais da organização.

Segundo William H. Inmon [11], os DWs não são construídos de uma só vez. Em vez disso, eles são projetados e povoados passo a passo, sendo, portanto, evolucionários e não revolucionários.

Um DW pode ser caracterizado como uma coleção de dados orientada por assunto, integrada, não-volátil, variante no tempo, que dá apoio às decisões da administração [11]. Os DW se diferenciam dos bancos de dados usuais, que são transacionais, pois são direcionados, geralmente, para aplicações de suporte à tomada de decisões. Eles armazenam os dados históricos e se dedicam principalmente a recuperação desses dados e não a processamento frequente de transações.

A Figura 3.1 mostra o DW como item central na solução de um BI recebendo dados de várias fontes de dados como bases de dados, arquivos e planilhas, e servindo de consulta para os relatórios.

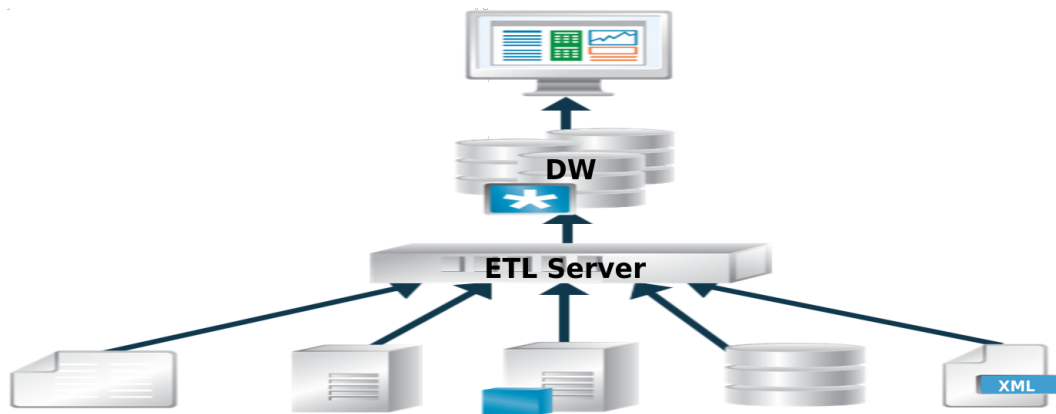


Figura 3.1: DW como Item Central na Solução de um BI [31].

Com relação à modelagem, os DW se utilizam de modelos diferentes dos usualmente conhecidos como relacionais. Ele se utiliza de uma modelagem específica para esse tipo de serviço, a modelagem multidimensional. Além disso, outra característica é que ele geralmente armazena grandes quantidade de informações organizacionais e, por isso, podem receber carga de múltiplos bancos de dados de diferentes tipos de modelagem. Para receber esses dados de outras origens é preciso, primeiramente, fazer uma limpeza e re-formatação desses dados antes de serem colocados dentro do DW, para que se adequem ao modelo multidimensional.

Como o intuito principal do DW é ter um banco que dê suporte às informações gerenciais de uma organização ao longo de sua existência, é muito importante que seus dados sejam armazenados com um histórico muito maior do que os bancos de dados transacionais, geralmente, costumam guardar. É por isso que ocorre apenas a inclusão de novos registros e não há atualização ou exclusão dos dados armazenados, a não ser que dados tenham sido inseridos incorretamente. O grau de usabilidade é muito importante no DW, já que o processamento pode se tornar muito complexo, devido à enorme quantidade de parâmetros, como granularidade e diferentes tipos de cruzamentos de dados para obtenção de informações gerenciais.

Existem duas metodologias clássicas usadas no desenvolvimento de DWs. O *bottom-up* (processo de baixo para cima), desenvolvido por Ralph Kimball [15], no qual um Data Mart (DM) é criado, inicialmente, para fornecer relatórios e capacidades analíticas para processos de negócios específicos, resultado de uma primeira análise dos processos de negócio relevantes a serem modelados. O DM é um subgrupo de um DW que atende a uma área específica de negócio da organização (algo similar a um DW específico de um departamento) e pode ser desenvolvido e vinculado de acordo com o ritmo de expansão do armazém de dados [2]. Os DMs podem eventualmente serem integrados para criar um DW maior.

Por outro lado, na metodologia *top down* (processo de cima para baixo), desenvolvida por William H. Inmon [11], inicialmente é criado um DW para que depois sejam definidos os DMs dimensionais de dados que comportam os dados necessários para processos de negócios específicos.

3.2.1 Sistemas Transacionais e Sistemas Analíticos

Efetivamente os sistemas transacionais não são projetados para produzirem informações estratégicas, o que torna esses sistemas inapropriados para o apoio à tomada de decisões [7]. Os sistemas *Online Transaction Process* (OLTP), ou transacionais, têm foco na rotina dos processos sem preocupações de mantimento de histórico e relatórios forjados para atender às demandas dos seus operadores, tendo assim pouca flexibilidade e difícil extração, pois não permitem que os gestores realizem análise e cruzamentos com outras informações, características essenciais para um armazém de dados.

Com as ferramentas *Online Analytical Processing* (OLAP) é possível diminuir a preocupação com a forma que os dados estão organizados e aumentar a atenção às informações contextualizadas que esses dados trazem ao gestor, melhorando o processo de tomada de decisão. Na Tabela 3.1, podem ser observadas as diferenças entre OLTP e OLAP.

Tabela 3.1: Diferenças entre os Sistemas OLTP e OLAP.

Características	OLTP	OLAP
Função	Automatizar operações diárias	Auxiliar na tomada de decisão
Usuário Humano	Cliente Atendente	Executivo Analista Engenheiro de conhecimento
Modelo Lógico	Relacional	Multidimensional
Granularidade	Única Atômica	Múltipla Agregada
Temporalidade dos Dados	Apenas valor corrente Atualizada constantemente	Histórico dos valores Completada periodicamente
Consultas	Simples Pré definidas	Complexas Ad-Hoc
Direção	Ler e escrever	Essencialmente ler

Enquanto os OLTPs tem o tempo de resposta das transações como um requisito crítico, e influenciando diretamente na agilidade e na qualidade do serviço prestado pelo sistema onde segundos fazem a diferença, os SADs podem ter respostas em minutos, horas e até dias, pois sua prioridade é a agregação de grande volume de dados para geração de relatórios para o desdobramento de uma informação gerencial ou até mesmo estratégica da organização.

3.3 Modelo Multidimensional

O processo de modelagem de um DW é composto por quatro passos: seleção do processo de negócio (o fato), definição do grão (o componente atômico do sistema), escolha das dimensões que se relacionarão com o fato e a identificação das métricas [16].

Os modelos multidimensionais utilizam os chamados cubos de dados para estruturar seus dados em matrizes multidimensionais. Para dados que se prestam à formatação dimensional, o desempenho de matrizes em consultas multidimensionais pode ser muito melhor do que no modelo de dados relacional [15]. O modelo tridimensional, por exemplo, que forma o cubo de dados (acima de três dimensões o cubo é chamado de hipercubo), pode ser consultado diretamente em quaisquer combinações de dimensão, evitando que sua consulta fique tão complexa como no modelo de dados relacional, onde seria preciso fazer múltiplos *inner joins* (criação de uma nova tabela de resultados a partir da combinação dos valores de colunas de duas tabelas) para relacionar as tabelas desejadas.

O propósito do modelo multidimensional é alterar a organização dos dados para uma forma menos complexa do que a relacional, normalizada, que exige do usuário extremo conhecimento das regras dos relacionamento das tabelas para se fazer consultas. Por isso, a principal característica de um banco de dados forjado à modelagem dimensional é a uniformidade das consultas [15].

No modelo multidimensional alguns conceitos devem ser apresentados. O fato é o registro de um resultado previamente desconhecido, com origem na repetição de um evento. As métricas são atributos previamente desconhecidos de uma observação e são medidas apenas no momento em que ocorre o fato. A dimensão é uma estrutura que acumula dados com origem em eventos previamente conhecidos formando um conjunto de atributos correlacionados.

Existem dois tipos de tabelas na modelagem de dados multidimensional: a dimensão e a fato. A tabela dimensão consiste nas tuplas de atributos que uma dimensão possui. A tabela fato é constituída de valores quantitativos e numéricos, e contém um ponteiro para cada fato a ser representado, apontando para a sua respectiva tabela de dimensão, ou seja, agrupa todas as referências para os atributos necessários [2].

As chaves delegadas (*Surrogate Keys* - SKs) são chaves pertencentes à entidades de DW que identificam registros das tabelas fato e dimensão, e com possibilidade de haver diversas fontes externas de dados, existem casos em que uma informação precisa de identificadores diferentes. Por meio de uma chave delegada é possível garantir unicidade dos identificadores em todo armazém de dados. Em regra, as SKs são números inteiros e, por tal motivo, ocupam pouco espaço e permitem junções rápidas [2]. A tabela dimensão armazena chaves naturais (chaves primárias do sistema OLTP), e também chaves delegadas, ligando o modelo dimensional com o transacional. Na tabela de dimensão, as

chaves primárias realizam ligações com o modelo transacional e as chaves delegadas com a tabela fato.

A Figura 3.2 mostra como é representada, em um modelo multidimensional, a ligação entre um fato e as dimensões de um projeto de DW. Como pode ser observado na tabela venda (fato), ocorrem ligações com as tabelas produto, tempo, cliente e revenda (dimensões) por meio de chaves de delegadas. Cada uma destas chaves, dentro de uma tabela dimensão, é similar a uma chave primária e todas juntas, dentro da tabela fato, são similares a chaves estrangeiras. Nesse exemplo, o modelo parece ser equivalente a um modelo relacional, porém não é. Ao contrário de uma chave primária comum, a chave delegada identifica e preserva cada versão de um registro. Chaves delegadas, primárias e estrangeiras coexistem em DW.

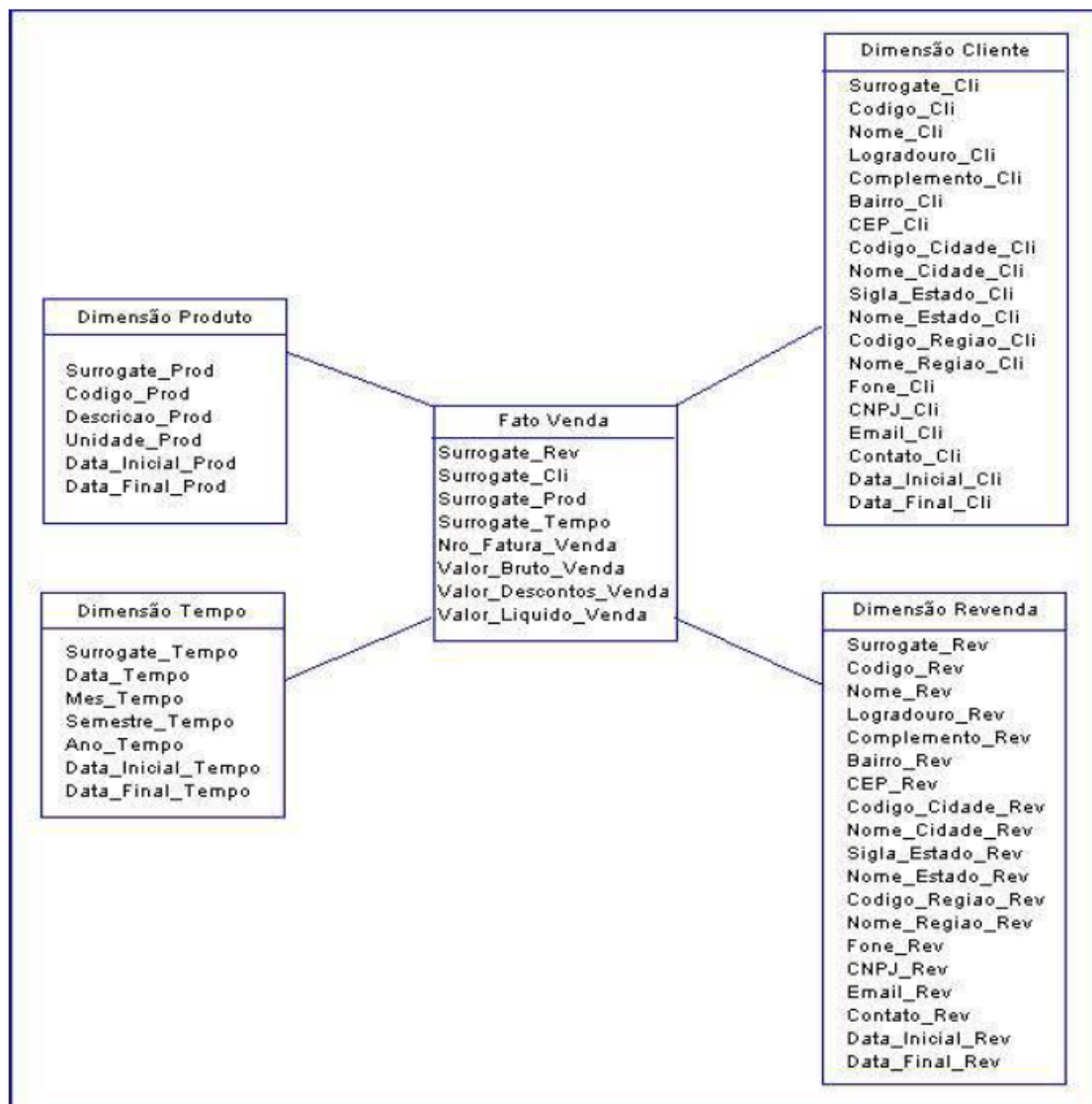


Figura 3.2: Modelo Multidimensional (Ligação entre o Fato e as Dimensões) [2].

3.3.1 Granularidade

Definir a granularidade é o processo de escolha do detalhe dos dados contidos no DW. O nível mais baixo de granularidade é designado de atômico, significando que não pode ser dividido. Escolher um nível de grão atômico é desejável, porque permite aos utilizadores agregar os dados à sua vontade, enquanto que escolher um nível intermédio de grão implica o risco de não ser possível satisfazer todas as pesquisas solicitadas pelos utilizadores. Segundo Ralph Kimball [16], é preferenciável desenvolver modelos dimensionais para a informação mais atômica capturada pelo processo de negócios.

Uma das primeiras decisões feitas para modelar o DW está relacionada com o nível de detalhe das métricas a serem armazenadas. Este nível de detalhamento é conhecido como granularidade da tabela fato. É muito importante que todas as linhas na tabela fato sejam armazenadas com informações exatamente no mesmo nível de detalhes. Como exemplo, um processo de pedidos teria sua granularidade definida no nível de detalhe da linha individual do pedido. Se forem armazenados diferentes níveis de detalhamento na tabela fato, a utilização da base de informações ficaria bastante prejudicada. Suponha que a maioria das informações no nível da linha de detalhe do pedido esteja armazenada, sendo que algumas informações foram armazenadas no nível do cabeçalho do pedido. No nível de detalhe da linha do pedido existe um relacionamento com cliente, produto, representante comercial e tempo. Porém, no nível do cabeçalho do pedido, o relacionamento com os produtos específicos por pedido desaparece. As informações relevantes que estejam em diferentes níveis de granularidade deveriam ser armazenadas em tabelas fato diferentes [12].

3.3.2 Dimensões

As tabelas dimensão contêm informações sobre as dimensões dos dados, ou seja, tempo, produto, mercado, contas, e, assim por diante, devem ser desenhadas a partir da perspectiva do usuário. Por esta razão, os atributos e suas descrições devem ser definidos de forma significativa para o usuário e adequados para a posterior exibição em relatórios.

A principal função da tabela dimensão é reunir os atributos que serão utilizados para qualificar as consultas e cujos valores serão utilizados para agrupar e sumarizar as métricas. Cada tabela dimensão deve ter múltiplos atributos que contenham valores ou textos que possam ajudar a descrever a chave.

As colunas dos atributos são utilizadas para filtrar o conteúdo da dimensão. Além disso, a utilização de atributos do tipo inteiro, quando apropriados, favorecem as operações de filtragem como maior que, menor que e entre outras [12].

3.3.3 Métricas

Uma métrica é aditiva quando é relevante somar seus valores ao longo de qualquer dimensão presente no sistema, ou seja, são eficazes, pois não possuem restrições de uso. Podem ser somadas em uma combinação qualquer de valores seja qual for a dimensão. Ao contrário da métrica aditiva, as métricas não aditivas não podem ser somadas ao longo dos valores das dimensões.

No caso de um sistema de vendas específico, por exemplo, existem duas métricas, as quais são margem de lucro e valor do pedido, que podem ser expressas como uma

porcentagem de vendas por meio da métrica margem percentual. As métricas margem de lucro e valor do pedido são aditivas, pois são expressas por somas de valores. No entanto, a métrica margem percentual é classificada como não aditiva porque é resultado de uma operação fracionária. Essa métrica poderia ser incluída na tabela fato, porém não seria útil. Portanto, seria mais eficaz armazenar na tabela fato apenas as métricas aditivas porque o cálculo que expressa a margem percentual pode ser feito a partir da divisão entre margem de lucro e o valor do pedido [12].

O terceiro tipo de métrica é a semi-aditiva que permite a soma ao longo de determinadas dimensões, porém não é aditiva para as dimensões restantes do sistema.

No exemplo de um sistema que gerencia o saldo das contas de um banco. Cada cliente pode ter acesso ao saldo armazenado no final de cada dia, por conta, ao longo do tempo. No caso do cliente possuir duas contas (corrente e poupança), ele pode somar os saldos de cada conta após o término do dia. A agência do cliente também pode fazer o somatório do saldo de todas as contas e obter um panorama geral da situação na localidade. Entretanto, não há utilidade no sentido adicionar o saldo de um cliente ou banco ao decorrer do tempo. Nos dois casos anteriores, o saldo funciona como uma métrica aditiva. Por esta razão, a métrica saldo é considerada semi-aditiva [12].

3.4 *Extract, Transform and Load (ETL)*

Segundo William H. Inmon [11], dados primitivos são dados detalhados utilizados na condução das operações cotidianas da empresa. Dados derivados são resumidos ou calculados de forma a atender às necessidades da gerência da empresa.

O ETL é o processo de extrair de diversas fontes de dados, que podem ser desde arquivos tabulados, planilhas e até base de dados, transformando estes dados para serem carregados, geralmente, em uma única base de dados, o DW.

Não há nenhum razão para se buscar dados do ambiente operacional para o ambiente de DW, sem integrá-los. Se os dados chegam ao armazém de dados num estado não integrado, não podem ser usados para suportar uma visão corporativa que é uma das essências de um ambiente arquitetado para funcionar dentro da inteligência de negócios [11].

Em cada ambiente, os dados operacionais não integrados são complexos e difíceis de serem utilizados. Essa é apenas uma realidade de qualquer sistema baseado em DW. E a tarefa de fazer essa integração de maneira mecânica e repetitiva não é um procedimento simples. Então para alcançar os benefícios reais que um armazém de dados pode oferecer, no entanto, é necessário se submeter a esse exercício difícil, complexo e demorado. O software ETL pode automatizar grande parte do processo de integração de dados. Além disso, é necessário que este processo de integração possa ser feito apenas uma vez. Mas, em qualquer caso, é obrigatório que o fluxo de dados para o DW mantenha-se íntegro e não fique alheio dentro do armazém de dados do ambiente operacional [11].

A área de estagiamento de dados (*staging area*) é uma área intermediária que funciona como uma área de armazenamento de um conjunto de processos em comum referidos como dados em processo de ETL. A área de estagiamento de dados é a ligação entre a base dos sistemas operacionais e a área onde os dados estão estagiados [15].

A extração é a etapa onde os dados são selecionados e depois alocados na área de estagiamento. É bem provável que no meio dos dados selecionados existam dados não

integros, havendo a necessidade de um tratamento para que sejam padronizados. Consequentemente, a partir de um processo de limpeza e tratamento, ocorre a transformação desses dados para que obedecem um formato padrão e se tornem consistentes e integros. Por fim, no momento em que há garantia da consistência dos dados, eles são movidos para área de estagiamento e são carregados no DW. [30]

Por meio dessa dinâmica que acontece a transformação em que os dados primitivos geram dados derivados que serão de extrema importância para os resultados apresentados [11]. A partir da criação de um modelo multidimensional dos dados, etapa necessária na implementação de um DW, é criado o processo ETL que vai executar o carregamento, do modelo projetado e as suas atualizações, em um banco de dados .

3.5 *Online Analytical Processing (OLAP)*

OLAP é um termo utilizado para se referenciar aos aplicativos feitos para consumir os dados do DW e criar uma interface com o usuário final, de maneira que ele possa interagir com o sistema, fazer consultas, montar relatórios e planilhas com granularidades e campos diversos, criados de acordo com seu interesse e visão de negócio, dando apoio às tomadas de decisão necessárias a diversas situações encontradas dentro de uma organização.

O processamento e a análise dos dados providos pelo DW é o momento crucial para que o gestor possa começar a usufruir dos frutos oferecidos pelo conceito de BI. É a partir do OLAP que uma solução de BI começa a chegar no seu usuário final. A abstração de todos os dados técnicos já está em um nível mais alto, o que proporciona ao usuário fazer suas consultas de maneira dinâmica e intuitiva sem que haja um nível de instrução técnica de TI para utilizá-lo. As principais funcionalidades de uma ferramentas OLAP são [2]:

- *drill-down*: navega entre as hierarquias do sistema, expandindo o nível de detalhamento de um grupo;
- *roll-up*: generaliza dados, classificando-os em grupos maiores (operação inversa ao *drill-down*);
- *slice*: seleciona, para a avaliação, um membro ou grupo de membros específicos de uma dimensão;
- *dice*: relaciona membros de um dimensão com os membros de outra dimensão;
- *drill-across*: inverte a posição entre duas dimensões relacionadas;
- *pivoting*: troca de eixos entre duas dimensões relacionadas;
- *ranking*: classifica os dados com medidas numéricas em ordem crescente ou decrescente;

A Tabela 3.2 mostra o resultado de uma operação de *slice*. Como pode ser observado, a partir da seleção de dois membros (produtos B e D) da dimensão produto e o confronto com o fato venda, foi gerada a tabela volume de vendas, que mostra a quantidade de vendas de cada membro e a soma da quantidade total de vendas dos dois membros.

A Tabela 3.3 mostra o resultado de uma operação de *dice* entre as dimensões revenda e tempo. A partir do confronto das dimensões revenda e tempo foi gerada a tabela

Tabela 3.2: *Slice* por Produto [2].

Volume de Vendas	
Produto	Total de Vendas
Produto B	2200
Produto D	812
Total	3012

Volume de Vendas, que mostra a quantidade de vendas dos membros das dimensões Revenda 1, 2, 3 e 4 ao longo dos meses Janeiro, Fevereiro, Março do ano de 2006.

Tabela 3.3: *Dice* entre Revenda e Tempo [2].

Volume de Vendas			
Revenda	2006		
	Janeiro	Fevereiro	Março
Revenda 1	250	300	280
Revenda 2	280	310	305
Revenda 3	212	220	231
Revenda 4	300	340	330
Total	1042	1170	1146

A Tabela 3.4 mostra o resultado de uma operação de *drill-across* entre as dimensões revenda e produto. A partir da troca de posição entre as dimensões revenda e produto, foi gerada a tabela Volume de Vendas - Revenda 1, que mostra a quantidade de vendas, das dimensões Produto A, B, C e D, no membro revenda 1, ao longo dos meses Janeiro, Fevereiro, Março do ano de 2006.

Tabela 3.4: *Drill-Across* entre Revenda e Produto [2].

Volume de Vendas - Revenda 1			
Revenda	2006		
	Janeiro	Fevereiro	Março
Produto A	75	85	81
Produto B	65	78	72
Produto C	59	73	80
Produto D	51	64	47
Total	250	300	280

Na Seção 3.4, foram abordados os conceitos básicos (granularidade, dimensão e métricas) essenciais para a construção, por meio de uma ferramenta OLAP, de um cubo. Um cubo é formado por métricas e dimensões e, no modelo multidimensional, é representado

por uma tabela fato. Uma métrica pode representar uma fórmula presente em um cubo ou o próprio cubo. A dimensão é formada por atributos e, no modelo multidimensional, tem a sua representação em uma tabela própria. Uma dimensão, normalmente, é formada por grupos hierárquicos, cada qual representa um subconjunto, para organizar os atributos em níveis. Um nível é representado por um atributo da dimensão que se relaciona com os demais atributos para garantir a hierarquia. Por fim, o atributo representa uma propriedade que se encontra na base da hierarquia de uma dimensão.

A Figura 3.3 mostra o projeto de um cubo tridimensional, onde cada eixo do cubo representa uma dimensão que está ligada a um fato.

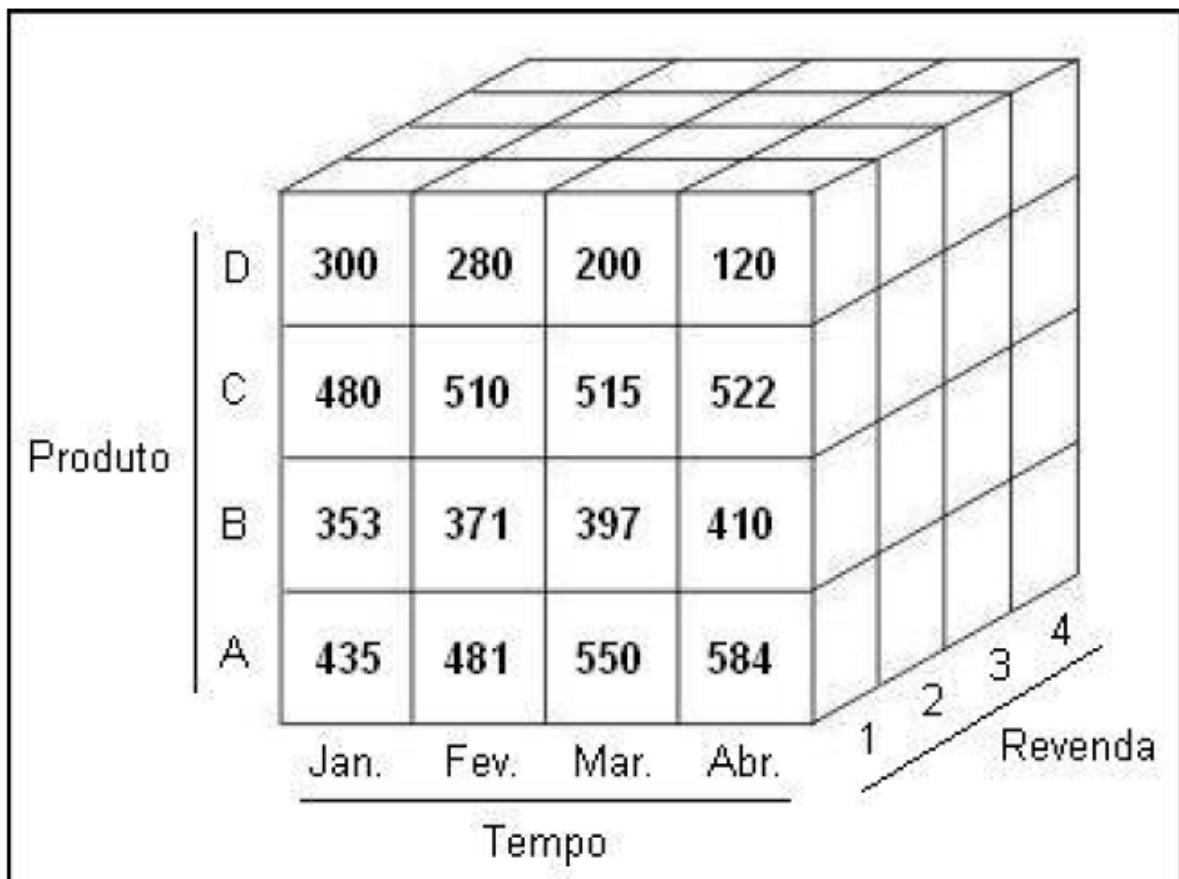


Figura 3.3: Cubo OLAP Tridimensional [2].

Capítulo 4

Plataforma Pentaho

O Capítulo 4 apresenta a descrição de todos os componentes tecnológicos utilizados na implementação do projeto, sendo dividido nas seguintes seções: a Seção 4.1 apresenta a visão geral da plataforma Pentaho, a Seção 4.2 apresenta o processo de integração dos dados no Pentaho, a Seção 4.3 apresenta as características de um servidor BI, a Seção 4.4 apresenta as ferramentas Modrian, MDX e *Pentaho Schema Workbench* (PSW), e a seção 4.5 apresenta a ferramenta *Pentaho Metadata Editor* (PME).

4.1 Visão Geral do Pentaho

Para a implementação deste sistema de informação foi adotado o conjunto de componentes do Pentaho devido a robustez, a simplicidade de uso e por ser software livre.

O Pentaho Possui uma arquitetura voltada para a implementação de software com suporte em BI. Baseada em tecnologias Java, a plataforma Pentaho pode ser usada em quaisquer sistemas operacionais e conectar em todos os bancos de dados que possuem *drives Java Database Connectivity* (JDBC). Os sistemas operacionais da Microsoft também são compatíveis com *Open Data Base Connectivity* (ODBC).

Na Figura 4.1 é mostrado o *Workflow* da Plataforma Pentaho. Cada etapa apresentada na linguagem é descrita a seguir. Os quatro passos mostrados na Figura 4.1 correspondem à:

1. Criação de *Data Warehouse*, *Data Mart* ou *dump* do banco com o *Pentaho Data Integration* (PDI), a partir de fontes de dados que podem ser bancos relacionais, serviços de rede, páginas web e fontes desestruturadas (como *e-mail* e documentos de texto), além de arquivos planos (csv, Excel, etc). Esse processo é conhecido como ETL;
2. Criação das soluções iniciais para exploração do DW: cubos *Online Analytical Process* (OLAP) usando o *Pentaho Schema Workbench* (PSW) [22], relatórios com ou sem parâmetros através do *Pentaho Report Design* (PRD) e *Web Ad-Hoc Query Reporter* (WAQR) com o *Pentaho Metadata Editor* (PME) e outros clientes de desenvolvimento podem ser usados (PDI, PDS e Weka);
3. Disponibilização da solução, ou seja, os relatórios e os resultados por meio do BI Server também conhecido por *Pentaho User Console* (PUC) e gestão de acesso

e permissões de visualização granulares inclusive com definições de grupos com o *Pentaho Administrator Console* (PAC);

4. Visualização da solução que pode ser via web por seus clientes ou até mesmo alguns clientes podem ter demandas especiais e optar por usar algum dos clientes de desenvolvimento, como PDI [20], Weka [25] ou *Pentaho Report Designer* (PRD) [24] para atendê-las.

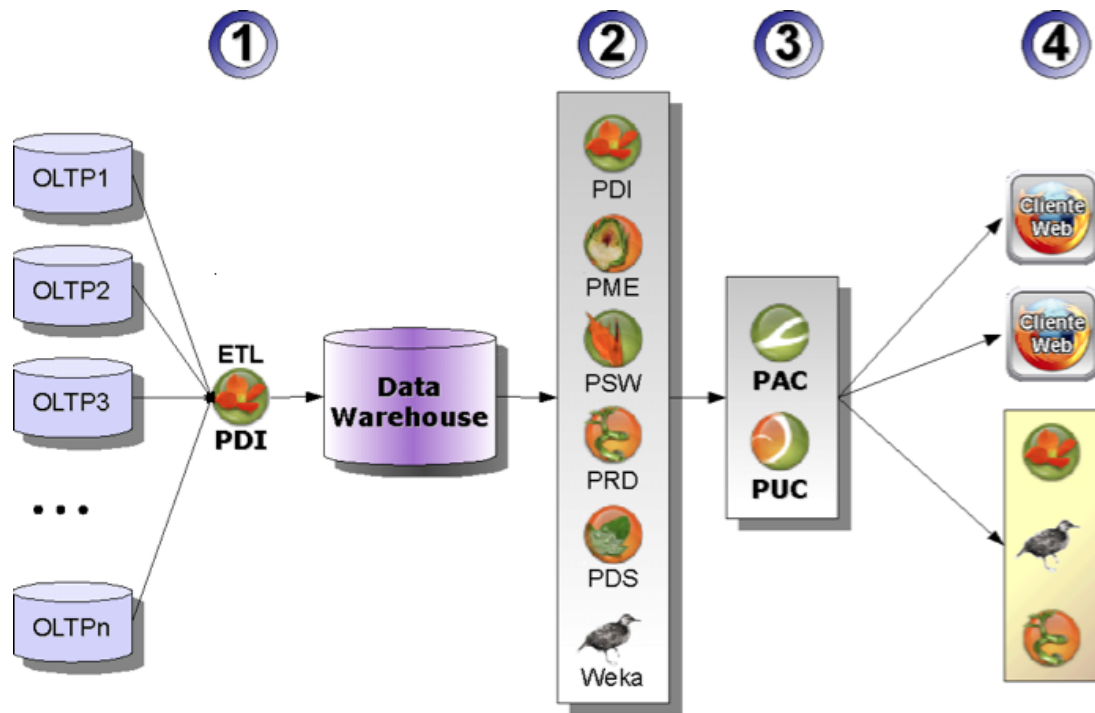


Figura 4.1: O *Workflow* da Plataforma Pentaho.

Outras vantagens desta plataforma de solução em BI são listadas na Tabelas 4.1.

Tabela 4.1: Vantagens da Plataforma Pentaho [33].

Customização	A plataforma de BI tem processos editáveis, disponibilidade de códigos fontes, componentes reutilizáveis e interfaces plenamente customizáveis. A plataforma tem mecanismos que podem ser integrados com processos de negócios.
Orientação à Ferramentas	A plataforma de BI é centrada em processos e soluções. Ela não é projetada para ser um conjunto de ferramentas isolado, mas um sistema completamente integrado.

Auditoria	A plataforma de BI inclui processos de rastreamento e auditoria. Os históricos podem ser acessados, incluindo as ações tomadas, discussões, comentários e informações de suporte.
Extensão	O código fonte é fornecido. A infraestrutura, os componentes, os mecanismos e as interfaces podem ser facilmente modificados. A estrutura do produto é pública.
Usabilidade	A plataforma BI tem componentes prontos para serem executados. É de fácil integração com sistemas externos, e oferece interfaces de fácil customização. Utiliza padrões de maneira a facilitar a integração com outras ferramentas.
Foco em Relatórios e Análise	A plataforma de BI é baseada em <i>workflow</i> e direcionada a regras de negócios. Ela pode ser facilmente integrada aos processos de negócios e tem facilidades para a geração de relatórios. Relatórios de desempenho de processos e modelos podem ser gerados.
Influência em Processos	A plataforma de BI acompanha a entrega de informações e permite processos de negócios customizados serem iniciados pelos receptores da informação. A plataforma oferece facilidades automáticas para realizar comparações de dados que auxiliem nas tomadas de decisões.
Prototipagem	O licenciamento da plataforma permite que a prototipagem seja realizada sem grandes investimentos.
Flexibilidade da Licença	O licenciamento da plataforma de BI é flexível, pois a Plataforma Pentaho utiliza uma licença <i>Open Source</i> (livre). A ferramenta <i>Pentaho Business Analytics</i> (PBA), parte da plataforma, possui duas versões: a versão comercial <i>Enterprise Edition</i> (EE) e a versão livre <i>Community Edition</i> (CE).

4.2 Integração de Dados no Pentaho

Segundo a seção *Infocenter* na página oficial do Pentaho, *Pentaho Data Integration* (PDI) é uma solução em extração, transformação e carga (ETL) que utiliza uma inovadora aproximação entre *metadata* e *hardware*. Inclui, uma interface gráfica, de fácil utilização, desenvolvimento e *design* para construção de *jobs* e *transformations* o que re-

sulta em um desenvolvimento rápido, menor custo de manutenção, depuração interativa e desenvolvimento simplificado [20].

No PDI, pode ser criado dois tipos de processos: as *transformations* e os *jobs*. A *Transformation* é a execução dos passos ligados por *hops* (saltos). Cada passo tem uma função específica pra ler, manipular ou gravar dados. As Figuras 4.2, 4.3, 4.4 e 4.5 mostram todos os passos (*steps*) disponíveis na versão 4.3.0 do PDI, onde é possível perceber o poder dessa ferramenta de integrar sistemas e bases de dados. Essa ferramenta trabalha com *big datas*, consumo de *web services* e disponibilização de relatórios. Os *hops* tem a função de caminhos de dados e definem o fluxo dos dados de um passo a outro tendo a opção de definir diferentes caminhos caso ocorra erros, ou de distribuir registros caso um passo siga para dois ou mais outras passos.

O *job* executa as *transformations* em uma sequência pré determinada, podendo ter comandos de decisão e laços (*if*, *else* e *loops*) além de se comunicar com administradores humanos através de *e-mail* ou mensagens de *log*. A execução sequencial é importante pois é preciso primeiro garantir que as dimensões sejam atualizadas para depois gerar os fatos. Além disso, também tem a opção de serem executados por agendamentos de tarefas do sistemas operacionais Linux ou Unix, utilizando o comando *cron*, e em sistemas Microsoft Windows utilizando o comando *schedule*. Um exemplo de *job* será mostrado no próximo capítulo referente a implementação.

As principais qualidades da ferramenta PDI são:

- realizar a integração dos dados e processos de ETL que alimentam DWs;
- ler e escrever mais de trinta formatos de SGBD, entre os quais Oracle, PostgreSQL e SQLServer;
- importar arquivos de texto (csv ou fixo), excel e bases de dados inclusive ODBC (apenas em nos sistemas operacionais Microsoft Windows), serviços de rede, páginas *web* e fontes desestruturadas (como *e-mail* e documentos de texto);
- estabelecer conexões com várias fontes de dados simultâneas;
- executar sequencias de passos, modificação e carga de dados em um destino;
- criar camadas de programas que são substituídas por operações visuais.

4.3 Servidor BI

O Pentaho BI Server é uma plataforma JSP [14] que roda sobre um servidor de aplicações Java, entre os quais JBoss [13] e Apache Tomcat [32]. Pode ser dividida em duas partes:

- o *Solution Engine*, ou mecanismo de soluções, responsável pela execução e controle de soluções, e funciona sobre a base de uma máquina de *workflow* interna;
- o Portal, que é uma parte visível onde o cliente final pode ter acesso as soluções.

É responsável por toda a camada servidora da plataforma Pentaho, fornecendo arquitetura e infraestrutura necessárias para construção de uma solução de BI ou inteligência de negócio. O *framework* provê serviços como autenticação, *login*, auditoria, *web service* e

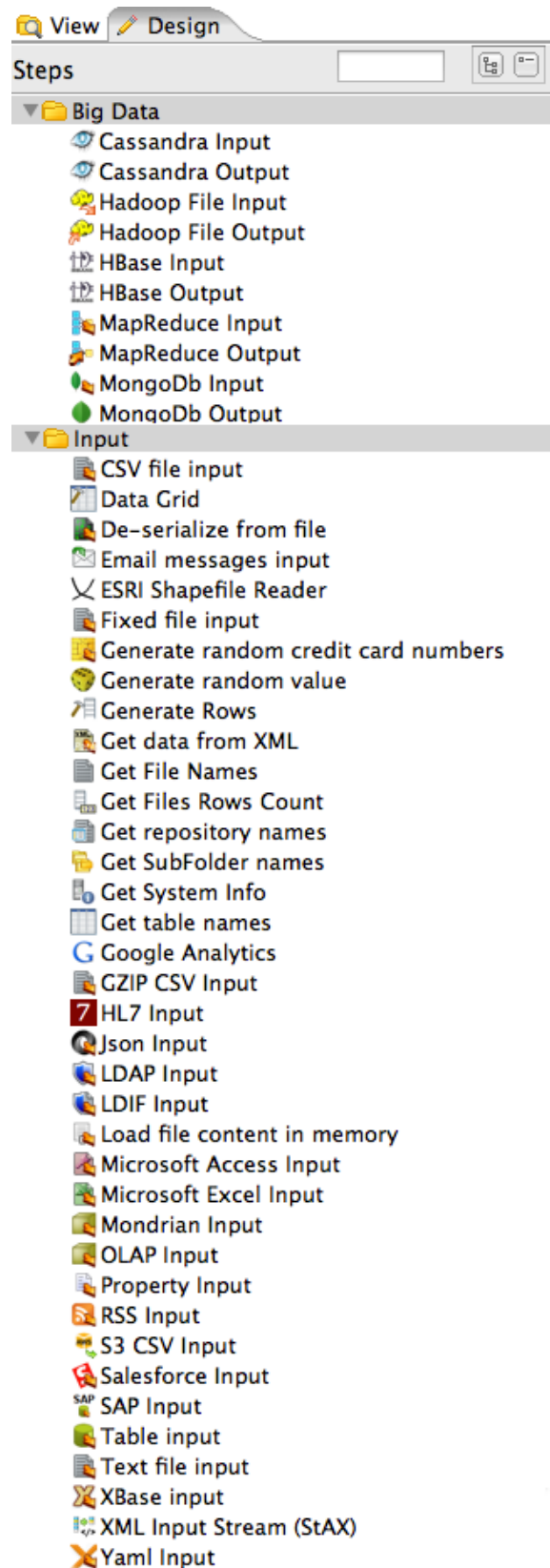


Figura 4.2: *Steps Big Data e Iput.*

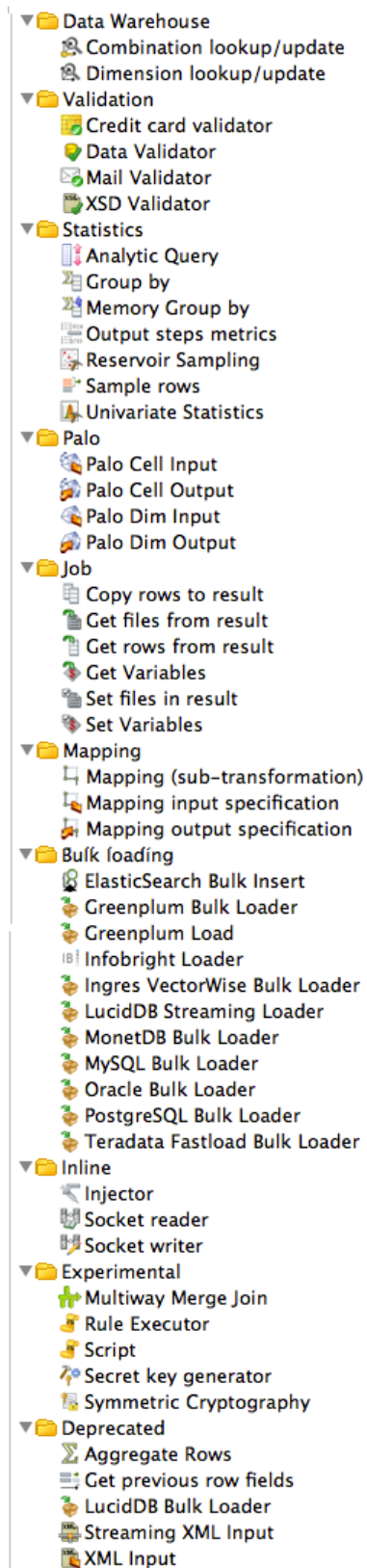


Figura 4.3: *Steps Data Warehouse, Validation, Statistics, Palo, Job, Mapping, Bulk Loading, Inline, Experimental e Deprecated.*

- ▼ Utility
 - Change file encoding
 - Clone row
 - Delay row
 - Execute a process
 - If field value is null
 - Mail
 - Metadata structure of stream
 - Null if...
 - Process files
 - Run SSH commands
 - Send message to Syslog
 - Write to log
- ▼ Flow
 - Abort
 - Append streams
 - Block this step until steps finish
 - Blocking Step
 - Detect empty stream
 - Dummy (do nothing)
 - ETL Metadata Injection
 - Filter rows
 - Identify last row in a stream
 - Java Filter
 - Prioritize streams
 - Single Threader
 - Switch / Case
- ▼ Scripting
 - Execute row SQL script
 - Execute SQL script
 - Formula
 - Modified Java Script Value
 - Regex Evaluation
 - User Defined Java Class
 - User Defined Java Expression
- ▼ Transform
 - Add a checksum
 - Add constants
 - Add sequence
 - Add value fields changing sequence
 - Add XML
 - Calculator
 - Closure Generator
 - Example plugin
 - Get ID from slave server
 - Number range
 - Replace in string
 - Row denormaliser
 - Row flattener
 - Row Normaliser
 - Select values
 - Set field value
 - Set field value to a constant
 - Sort rows
 - Split field to rows
 - Split Fields
 - String operations
 - Strings cut
 - Unique rows
 - Unique rows (HashSet)
 - Value Mapper
 - XSL Transformation

Figura 4.4: *Steps Utility, Flow, Scripting e Transform.*

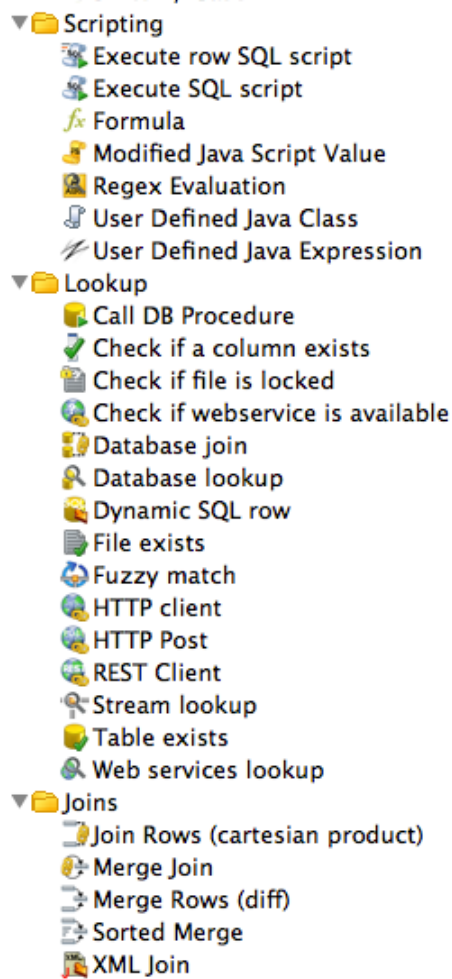


Figura 4.5: *Steps Scripting, Lookup e Joins.*

motores de regras de negócio. A plataforma também inclui disponibilização de relatórios, análises OLAP, mineração de dados e *dashboards* (painéis de indicadores).

É no BI *Server* PUC que o usuário final tem contado com os resultados da solução implementada, podendo acessar os relatórios já previamente disponibilizados ou usar serviços como OLAP ou *Web AdHoc Query Reporting* (WAQR), ou seja, escolher os dados e montar online seu próprio relatório a partir do DW com agrupamentos, filtros e *layouts* configuráveis, e com a possibilidade de visualização em tela ou exportar para os formatos excel, pdf, csv e salvar para futuras consultas e compartilhamento *online* no próprio BI Server. Todos esses serviços não exigem nenhum conhecimento em SQL ou as regras de estrutura do banco de dados do usuário final, geralmente, um gerente, e este podendo focar nas análises do processo.

O BI Server também provê uma seção de administração o PAC, onde se configura as conexões com as bases de dados, gerencia usuários e grupos de usuários e suas respectivas permissões de acesso, *jobs* ou rotinas a serem executadas.

O portal do cliente oferece serviços pré configurados como registro de soluções, controle de acesso, relatórios, agendamentos. Há outros serviços que podem ser construídos através do portal, entre os quais serviços de *dashboard* envio de *e-mails* programados.

A modularidade do portal permite que outros serviços sejam criados e implementados a critério do desenvolvedor.

O Pentaho BI Server utiliza como framework de segurança o Spring Security [29] para controle de autenticação forte e altamente personalizável e quadro de controle de acesso. O sistema de segurança do BI Server permite a criação de grupos de usuários e permissões granulares podendo ser atribuídas desde à dados para consulta, relatórios e à suas pastas.

4.4 Mondrian, MDX e *Pentaho Schema Workbench*

Segundo a documentação do Pentaho, Mondrian é uma *engine* para Processamento OLAP escrito em Java. Ele executa consultas escritas em linguagem MDX, lendo dados de um banco de dados relacional e apresentando os resultados em um formato multidimensional via uma API Java. Permitindo clientes analisem grande quantidades de dados em tempo real. MDX significa *multi-dimensional expressions*. É a linguagem de consulta principal da plataforma Pentaho [21].

MDX foi introduzido pela Microsoft com o *Microsoft SQL Server OLAP Services* em torno de 1998, como o componente de linguagem do *Object Linking and Embedding, Database* (OLE DB) para *OLAP Application Programming Interface* (OLAP API). Mais recentemente, a MDX apareceu como parte do XML for Analysis API. A Microsoft propôs MDX como um padrão, e sua adoção entre os autores de aplicativos e outros prestadores de OLAP é cada vez maior [21].

A plataforma Pentaho permite que o modelo MDX seja adaptado e traduzido para a realidade da ferramenta, através da tradução da sintaxe do modelo SQL da Microsoft.

Para que o Mondrian possa ler banco de dados de qualquer tipo e fornecer um cubo OLAP, ele precisa saber como se mapeiam os elementos de um no outro. O esquema Mondrian, um arquivo XML com uma estrutura bem definida, é quem fornece esse mapeamento lógico da base de dados multidimensional presente nos cubos como dimensões, hierarquia, níveis e membros. Por tanto, o Mondrian é um *Relational Online Analytical Processing* (ROLAP), pois ele constrói seus resultados a partir de um banco de dados relacional.

A confecção desse esquema pode ser feita utilizando a ferramenta *Pentaho Schema Workbench* (PSW). O *Workbench* é um ferramenta que facilita o desenvolvimento do esquema, porém ainda não implementa todos os recursos avançados que o Mondrian é capaz de produzir e estes podem ser utilizados a partir da edição no próprio arquivo xml.

Então, a elaboração dos cubos OLAP se dá no PSW que gera o esquema este é publicado no BI Server, interface que permite as consulta por meio da linguagem MDX. Estas consultas são processadas pela *engine* Mondrian que finalmente retorna os resultados em tempo real.

4.5 *Pentaho Metadata Editor* (PME)

Segundo a seção wiki da página oficial do Pentaho, *Pentaho Metadata Editor* (PME) é uma interface gráfica de usuário para a criação de um modelo de metadados para usuário amigável em cima de fontes de dados relacionais. Esta visão de negócios permite aos administradores suprimir as complexidades das fontes de dados subjacentes, enriquecer

os modelos por meio da formatação, dados de supressão, descrições e aplicar a segurança funcional [23].

O PME, assim como o PDI, é uma ferramenta visual que possui elementos de interface gráfica similares. As suas principais qualidades são:

- criação de mapeamentos entre as tabelas que serão usadas pela plataforma para gerarem relatórios *Ad-Hoc* via web;
- criação de visão de negócios independentes, inclusive para campos, de visões distintas, que estão localizados em uma mesma tabela.

O PME também é utilizado para construir uma documentação e funciona como fonte de dados para o *Report Designer*. Como ferramenta de documentação permite que o desenvolvedor reconstrua modelos de entidade e relacionamento para bases de dados usadas como fonte para um DW, e através do modelo construído apresente uma opção para criação de relatórios.

Na Figura 4.6 são mostrados os elementos do metamodelo que foram organizados pela Pentaho numa ordem lógica e na sequência de criação. Assim, a primeira coisa a ser criada é a conexão com a fonte de dados e, posteriormente, as tabelas lógicas. Por fim as visões de negócio são criadas.

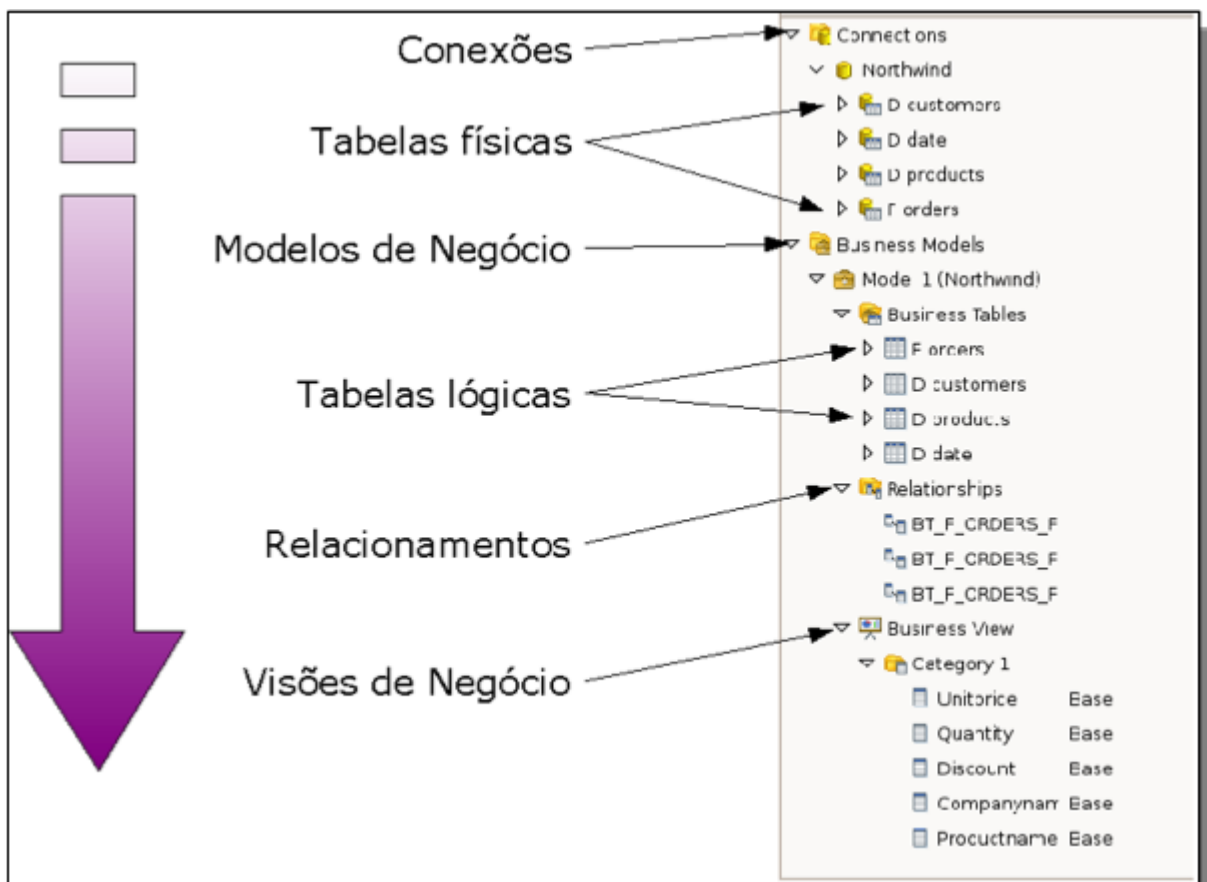


Figura 4.6: Processo de Criação de Metamodelos.

Capítulo 5

Implementação

Neste capítulo descreve-se como foi implementada a solução de BI, passo a passo, desde o diagnóstico do processo, a modelagem multidimensional, o ETL, a criação do DW, a criação do cubo OLAP MDX e a disponibilização dos relatórios no BI Server.

5.1 Análise (Diagnóstico) do Processo

No primeiro momento foi feito o levantamento da história dos sistemas usados para a gestão do processo de IC, que basicamente tem a função de manter o cadastro dos atores do processo que são os alunos, os orientadores e o controle dos projetos.

Nesse caso, foi constatado que houveram dois momentos. No primeiro, construiu-se um sistema em Access que era replicado anualmente em um arquivo distinto e essa sistemática ocorreu entre os anos de 2000 a 2003. No segundo momento foi desenvolvido o SibolWeb, sistema baseado em tecnologias web em linguagem PHP e base de dados MySQL de 2004 até o momento da escrita desta monografia.

Os editais ocorrem anualmente, sendo que cada projeto tem um orientador e pode ter até quatro alunos, cada um com o seu plano de trabalho. Os orientadores e alunos de um mesmo projeto podem ser de cursos e departamentos distintos.

5.2 Modelagem Multidimensional

Seguindo os passos apresentados na Seção 3.4, desta monografia, para a modelagem multidimensional, define-se o processo, a granularidade, as dimensões, as métricas e a tabela fato.

A partir do diagnóstico do processo de gestão da IC, por meio de entrevistas com os gestores e análise dos sistemas, foi possível identificar quatro dimensões e um fato. O fato neste momento são os planos de trabalhos desnormalizados com os projetos, e as dimensões são aluno, orientador, projeto e tempo.

Para a granularidade foi adotado o nível mais atômico, ou seja, o registro de cada plano de trabalho por aluno, orientador e ano do edital.

A métrica aditiva, chamada quantidade, é a ocorrência de cada fato, por isso é adotada uma constate de valor 1 em cada registro da tabela fato para que sua agregação seja uma

soma, facilitando o processamento, visto que, uma soma custa menos para o SGDB do que um contador produzido pelas operações *count* ou *count distinct* em SQL.

5.3 Processo de ETL

Para o processo de extração, transformação e carga na Plataforma Pentaho foi usado o programa PDI, onde são criadas *transformations* e *jobs* em uma interface gráfica. Foram realizadas separadamente dois processos de ETL, um para o Access e outro para o MySQL, que, conseqüentemente, foram exportados para o DW.

Os arquivos do Sibol estavam no formato mdb do Access e foram mapeados para a criação de um padrão comum entre as quatro versões de arquivos, relativos aos anos 2000, 2001, 2002 e 2003, sendo feita uma única transformação no processo ETL destes arquivos. Os dados do SibolWeb solicitados, referentes ao intervalo entre os anos 2004 e 2012, fazem parte de um arquivo único, no formato csv, de dados que estavam conforme a modelagem do DW, sem muitas alterações na disposição e formatação dos dados. Esta situação não elimina o processo de ETL, pois tem que se atribuir as chaves delegadas aos registros e relacioná-los na tabela fato.

Uma ferramenta para a gestão do ETL também foi utilizado chamada mapa lógico dos dados que contém todos os sistemas, bancos, campos, tipos em um formato que indica a transição da modelagem relacional para dimensional, facilitando também a integração dos dados, pois há uma variação entre bancos de dados, sendo necessário o estabelecimento de um único padrão no DW. Um exemplo é o campo sexo que assume valores como "m" ou "f", "masculino" ou "feminino", 0 ou 1, sendo neste caso definido como "M" ou "F". A Tabela 5.1 mostra o modelo de um Mapa Lógico dos Dados.

Tabela 5.1: Planilha Mapa Lógico dos Dados.

Gestão	Sistema de Origem Prioridade
Identificação no DW	DFS Tabelas Coluna Tipo SCD
Fontes	<i>Database</i> Tabela Coluna Tipo Tamanho Precisão
Transformação	Padrão

No passo de extração das bases, sejam elas csv ou Access, define-se os tipos e as máscaras dos dados. O campo IRA da tabela aluno é setada como número no formato *##.##* para que se comporte na carga como um tipo *float*. A Figura 5.1 mostra a tela de configuração de extração.

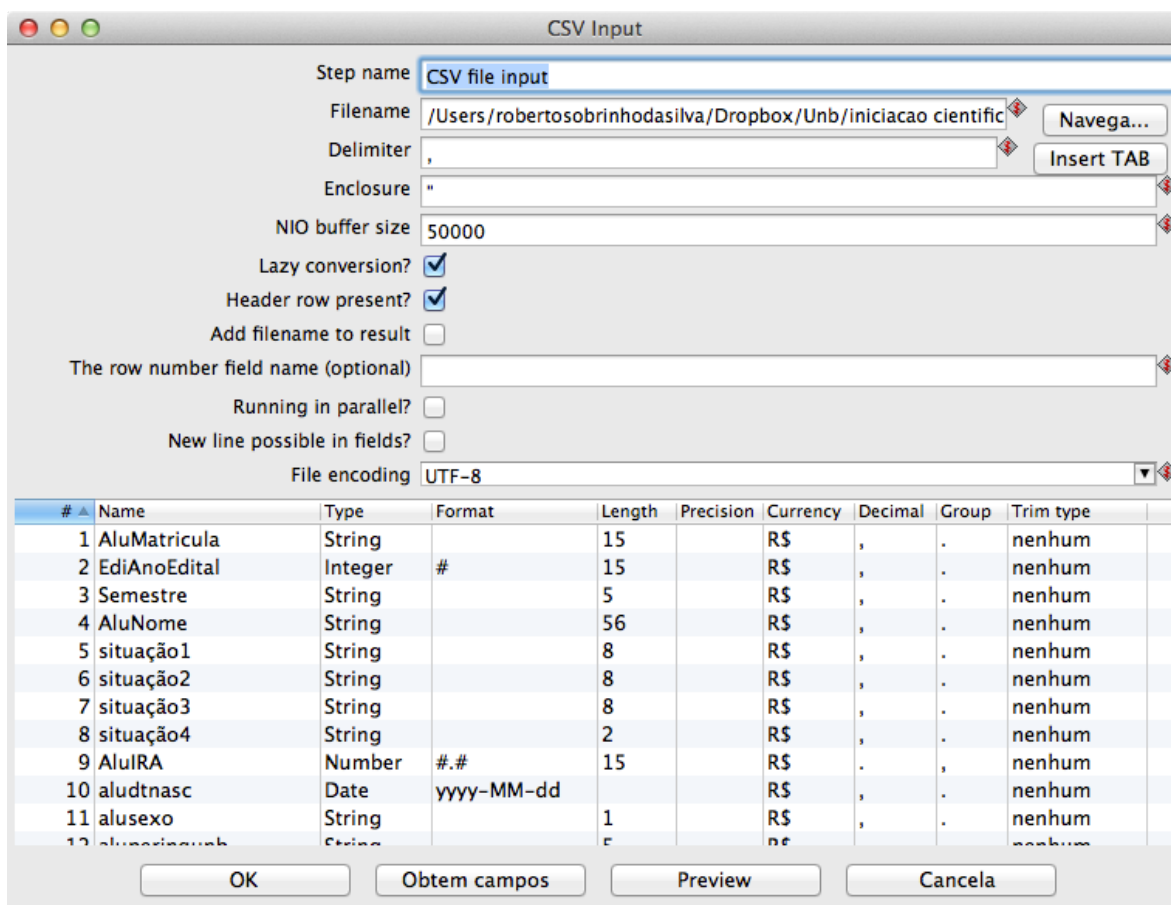


Figura 5.1: PDI - Extração.

Neste passo é muito importante a solicitação não apenas dos dados, mas também dos metadados, pois muitos valores estão abreviados em siglas ou índices que não são claros para tradução. Por exemplo, o tipo de bolsa do aluno que aparece, em alguns casos, com as denominações Remuneradas e Voluntária e em outros casos com as denominações RN, VL e NV.

Inicialmente, foram feitas as transformações dos arquivos tabulados, pois esses já estavam o mais próximo do padrão a ser adotado. O ETL do arquivo fornecido pelo CPD, atualizado em outubro de 2012, é um xml com 5 abas, sendo cada uma exportada para o um arquivo no formato csv.

Os dados dos alunos tiveram os campos selecionados, renomeados e carregados para o DW. Na aba orientadores foram eliminados os registros repetidos, sendo carregados apenas os registros únicos de cada orientador.

Os dados da aba projeto foram desnormalizados com os dados do plano de trabalho por meio das operações *order* e *merge join*. Outro recurso utilizado foi o mapeamento de valores, pois os campos relacionados com a área de conhecimento tinham valores nulos que foram substituídos pelo valor 0. A Figura 5.2 mostra essa transformação.

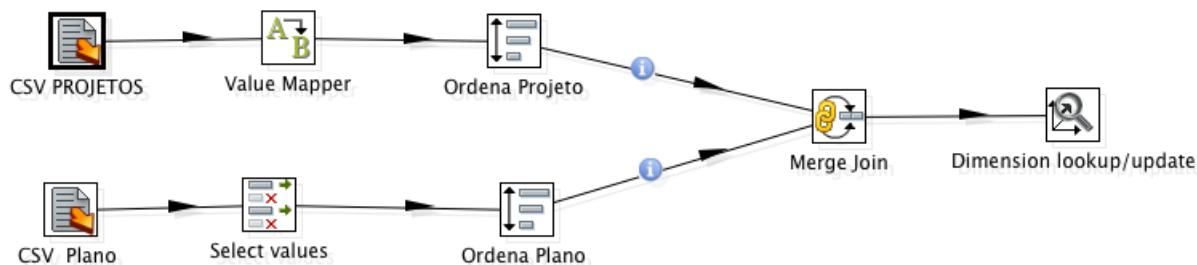


Figura 5.2: Transformação da Dimensão Projeto.

As transformações dos bancos de dados em Access precisam ter os campos agregados e os valores adequados para o padrão adotado porque são mais complexas, ou seja, exigem a implementação de mais passos. A figura 5.3 mostra as transformações da dimensão aluno. Assim, o processo ETL seguiu os seguintes passos:

- união das tabelas Aluno, Curso e Plano de Trabalho;
- carga dos registros da tabela Aluno;
- desnormalização da tabela curso;
- busca do campo tipo de bolsa na tabela Plano de Trabalho;
- criação de um *script* em java que aponta o ano do edital de acordo com o banco importado;
- transformação do CPF do tipo *string* para o tipo *int*;
- concatenação dos campos ano e semestre de ingresso do aluno, convertendo-os para o formato ano/semestre;
- mapeamento dos valores do campo sexo para o padrão adotado.

A criação da tabela fato é basicamente a consulta das últimas importações dos planos de trabalho através das chaves relacionais dos registros vinculados às dimensões aluno, orientador e projeto, retornando suas respectivas chaves delegadas e incluindo a métrica que, neste caso, é uma constante de valor 1 que por meio de uma soma, representará a quantidade de ocorrências do fato. A construção da tabela fato é similar nos ETLs dos arquivos csv e Access, e a única diferença entre as duas é o formato da fonte de dados. A Figura 5.4 apresenta os passos deste processo.

A dimensão data utiliza apenas o campo ano, portanto, é considerada uma dimensão degenerada. Para aumentar a performance das consultas, o campo ano foi agregado à tabela fato.

Nesta seção foram mostradas as etapas de extração e transformação dos dados. A carga é a etapa que fecha o processo de ETL, sendo descrita na próxima seção.

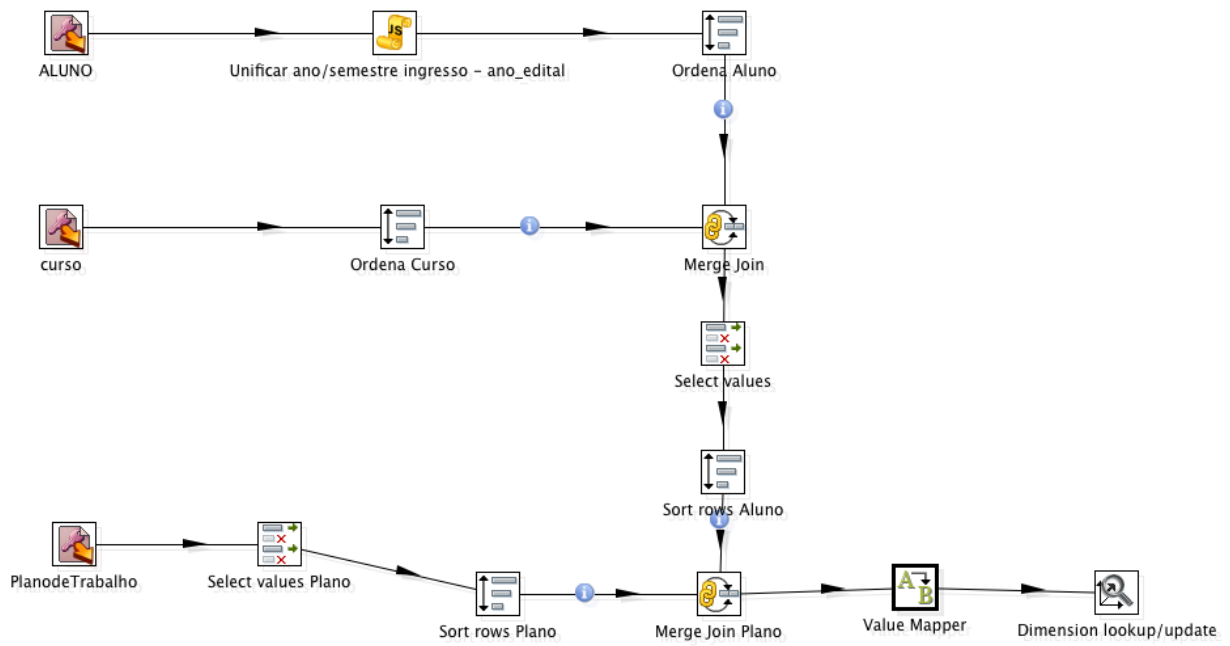


Figura 5.3: Transformação da Dimensão Aluno do Banco de Dados Access.

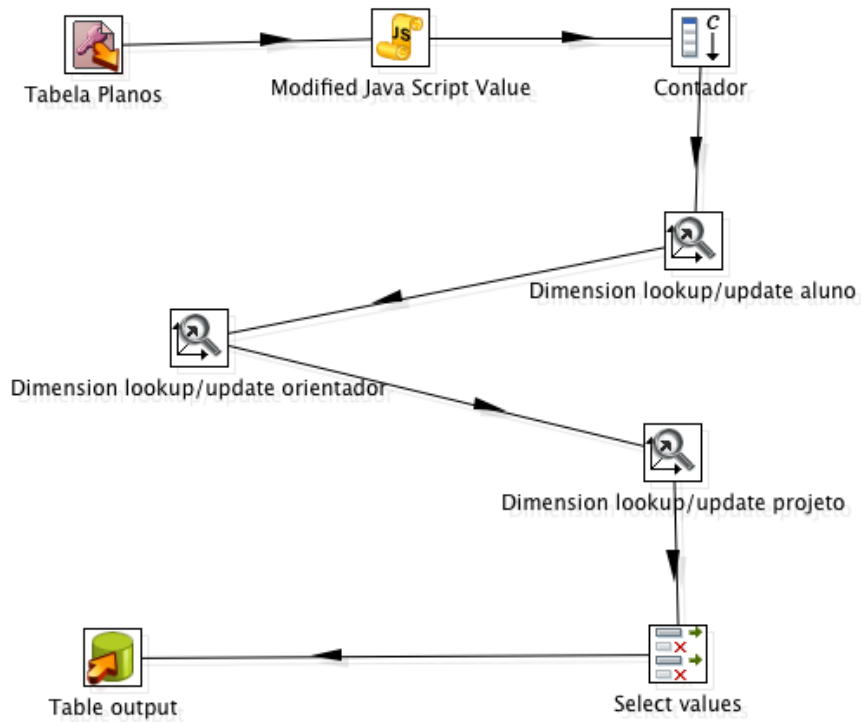


Figura 5.4: Transformação do Fato.

5.4 Criação do DW

O DW está sobre o SGBD PostgreSQL [8] que é um *software livre* de grande robustez e uma referência entre os SGBDs da sua categoria. Apesar de existirem alguns bancos de dados projetados para modelagem multidimensional, é bastante comum o uso de SGBDs relacionais para suportar projetos de DW.

A partir da modelagem foram criadas, utilizando o modelo multidimensional, as tabelas `dim_aluno`, `dim_orientador`, `dim_projeto` e `fato_projeto` em um banco de dados denominado `DW_PIBIC`.

Com a ferramenta PDI, no passo *Dimension lookup/update*, é possível obter os campos resultantes do ETL e as chaves do modelo relacional. Com as chaves do modelo relacional configuram-se a chave delegada, um campo para versionamento e as datas de validade do dado. O PDI se conecta no banco e sugere uma *query* em SQL para criação dos campos necessários para a realização da carga. Porém, geralmente é necessário uma adequação dos tipos e tamanhos destes campos e uma atenção com os comandos a serem executados, pois se existir algum campo no banco de dados que não está presente nesta carga ele fornece a opção de excluí-lo. Na Figura 5.5 é possível observar a tela de configuração deste passo do PDI e, ao lado, a *query* a ser executada.

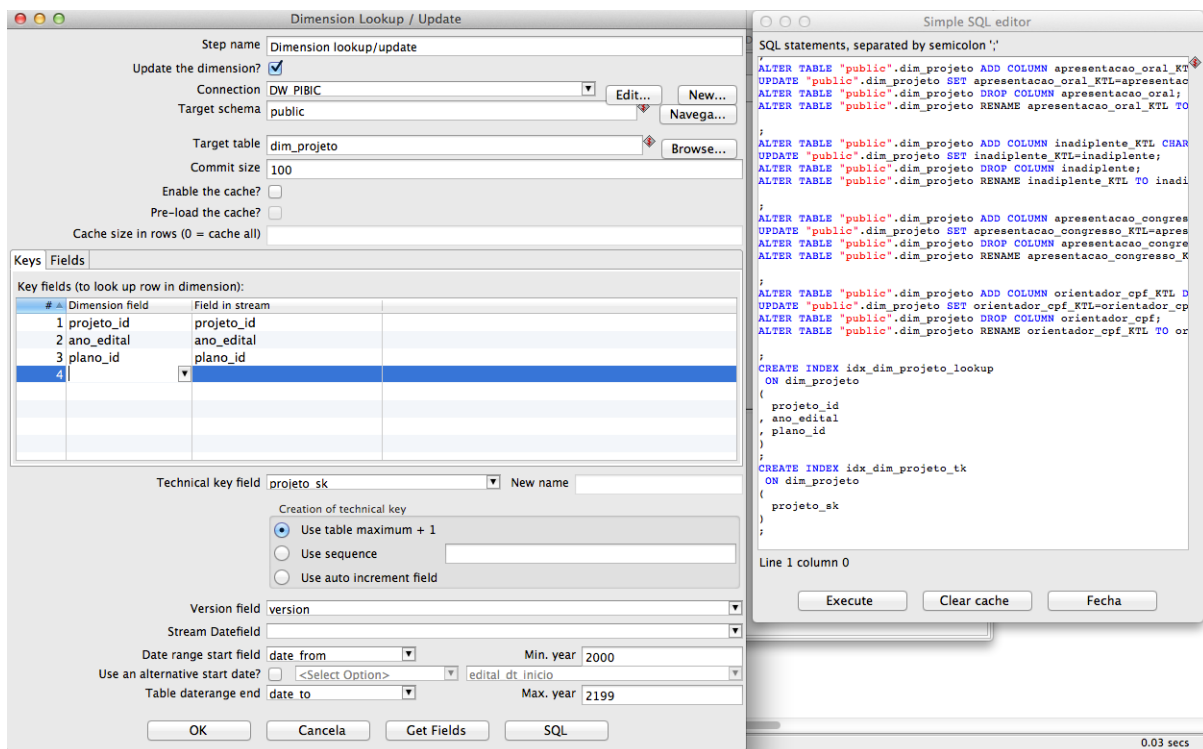


Figura 5.5: *Dimension Lookup/Update*.

5.5 Cubo OLAP MDX

O DW completo e consistente, com dados nos mesmos padrões e bem formatados, possibilita a criação do cubo OLAP para disponibilizar consultas multidimensionais. Para esta tarefa é utilizada a ferramenta PSW.

O objetivo é criar um esquema e, dentro deste esquema, um cubo. Cada cubo representa uma tabela fato com suas métricas e dimensões. No modelo dimensional adotado, o fato é o plano de trabalho, a métrica é a quantidade de ocorrências deste plano e as dimensões são aluno, orientador, projeto e ano.

Um esquema pode conter vários cubos e isso permite a criação de cubos específicos para o tipo de análise a ser realizada. O reuso das dimensões pode ser feito definindo-as como dimensões globais. A dimensão degenerada Ano do edital é definida como uma dimensão global, pois tem grande chances de ser reutilizada em novos cubos.

O membro calculado é o recurso utilizado para cálculos de campos usando a sintaxe da linguagem MDX . Estes cálculos podem ser feitos entre campos, inclusive campos de atributos de dimensões. Em uma implementação com este recurso calcula-se a média dos IRAs dos alunos vinculados aos projetos da iniciação científica.

Com todas essas definições é salvo o esquema no formato xml e em seguida publicado no BI Server para ser disponibilizado para análises, consultas, geração de relatórios e gráficos.

A Figura 5.6 apresenta o cubo vinculando à tabela fato_plano, e a sua métrica é a quantidade, ou seja, o somatório da constante de ocorrência do tipo inteiro. Na mesma figura tem-se um exemplo de hierarquia dentro da dimensão e neste caso a Lotação Aluno tem a hierarquia campus, departamento e curso.

5.6 Metadados e Relatórios WAQR

Após os processos de ETL e depois de o DW estar populado com os dados, é necessário disponibilizar os relatórios aos usuários finais e, para isso, existem várias opções. Uma delas é o cubo OLAP mostrado na seção anterior e outra forma é a geração de metamodelos para disponibilização de fontes de dados para relatórios WAQR ou parametrizados.

Seguindo o processo mostrado na Seção 4.5, elaborou-se o seguinte metamodelo mostrado na Figura 5.7, disponibilizado no BI Server, para ser usado como fonte para os relatórios WAQR.

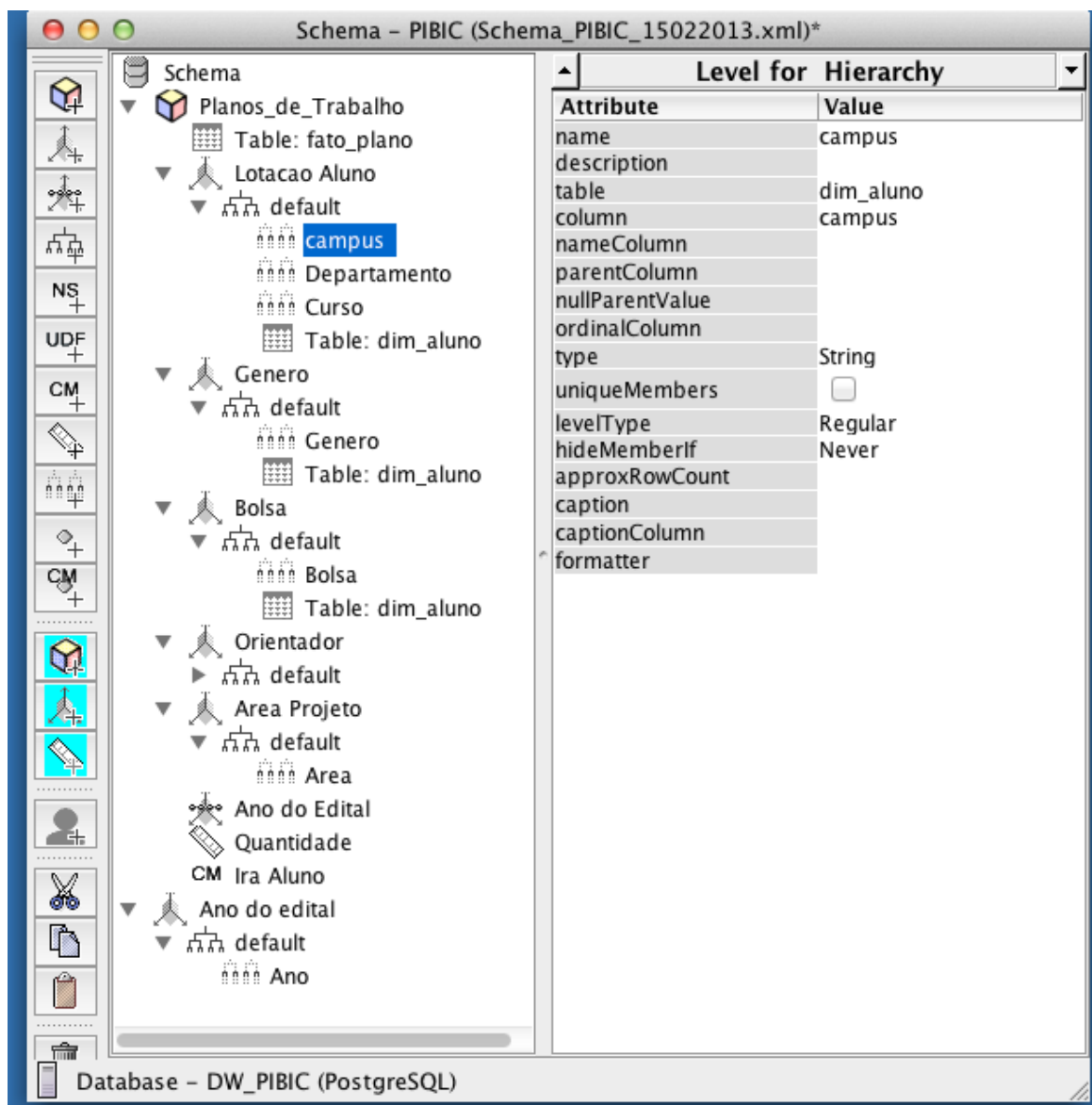


Figura 5.6: Cubo Vinculando a Tabela Fato_Plano.

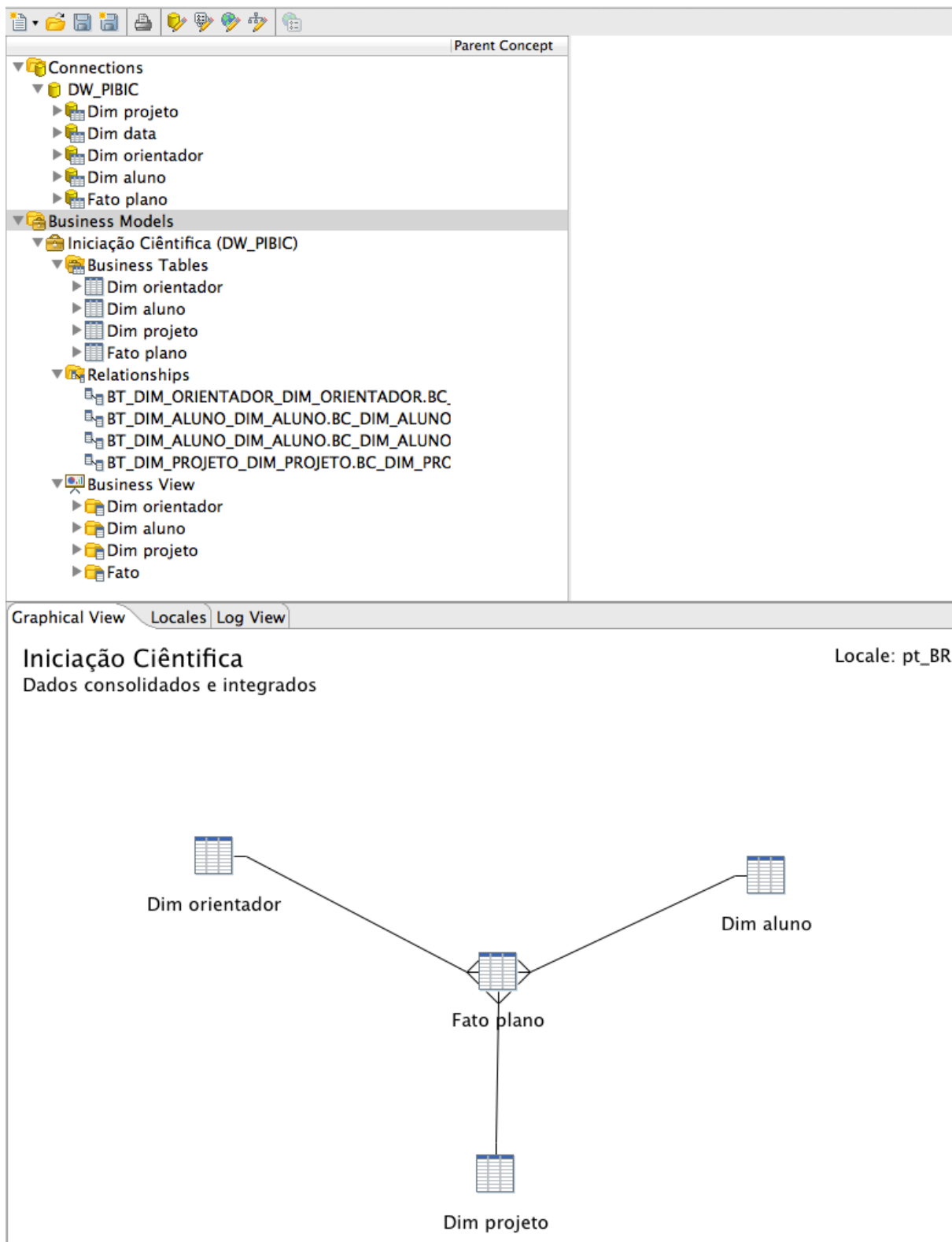


Figura 5.7: Fonte de Relatórios WAQR.

Capítulo 6

Resultados

No capítulo 6 são apresentados alguns relatórios que fazem parte dos resultados deste projeto. Esses relatórios foram baseados na solicitação feita pelo gestor da informação do PIBIC da UnB no período de desenvolvimento do projeto. Os relatórios apresentados são qualidade dos dados, perfil dos alunos, perfil dos orientadores, número de bolsistas no período e número de voluntários no período.

6.1 O Sistema de Informação Gerencial

O Sistema de Informação Gerencial para o PIBIC é uma customização do PUC, utilizando as marcas da UnB seguindo os padrões estabelecidos pela Secretaria de Comunicação da Universidade de Brasília [35]. As logos do rodapé com *links* foram retiradas da página do DPP [34]. A tela inicial de *login* é exibida na Figura 6.1.

A área de trabalho do PUC contém as opções para geração de novos relatório no centro da tela. A esquerda estão as pastas e relatórios salvos anteriormente organizados por pastas, como mostra na Figura 6.2.

A opção *New Report* remete à ferramenta de geração de relatórios WAQR com todos os campos disponíveis no DW agrupados em suas respectivas dimensões, podendo serem selecionados e agregados, a critério do usuário. Esta tela é mostrada na Figura 6.3, e na Figura 6.4 é mostrada a tela com o relatório obtido. Este relatório além ser visualizado em HTML, na própria tela, pode ser exportado para pdf, Excel, csv e ser salvo para consultas posteriores.

A opção *New Analysis* remete à consulta e navegação dos cubos OLAP, permitindo consultas dos dados do DW a partir da distribuição das dimensões e métricas entre linhas e colunas, aplicação de filtros, ordenação desejada e todas as operações de manipulação possíveis em um cubo. Está disponível também a geração de gráficos, as consultas em MDX, a visualização da tabela com a relação dos registros agregados e a exportação do resultado para Excel. Um exemplo de navegação em um cubo no PUC é mostrada na Figura 6.5, na qual é feita uma análise da quantidade de planos de trabalho do campus Ceilândia por curso, tipo de bolsa, graduação dos orientadores e ano do edital.



Sistema de Informação Gerencial para o Programa de Iniciação Científica (ProIC)

Alunos:

Roberto Sobrinho da Silva

Mario Matos Camargo Filho

Orientadora:

Dra. Maristela Terto de Holanda

Login



User Name:

Roberto

Password:

Launch in new window

Login



© 2005-2013 Pentaho Corporation. All rights reserved.

Figura 6.1: Tela de *Login*.

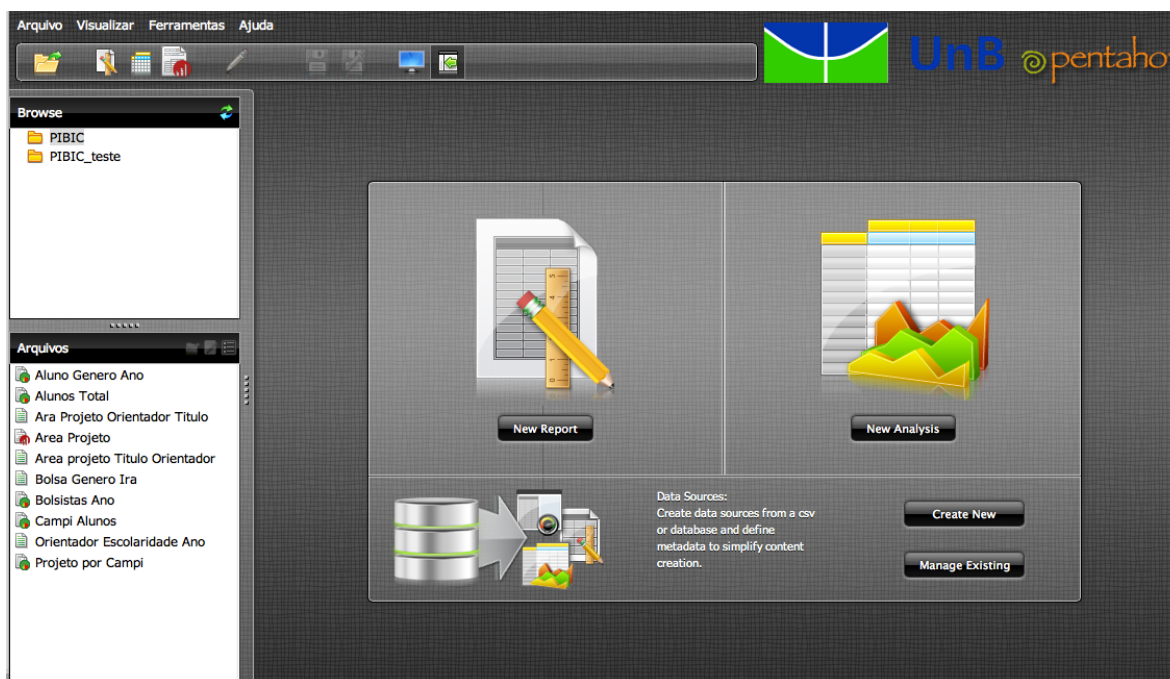


Figura 6.2: Área de Trabalho.

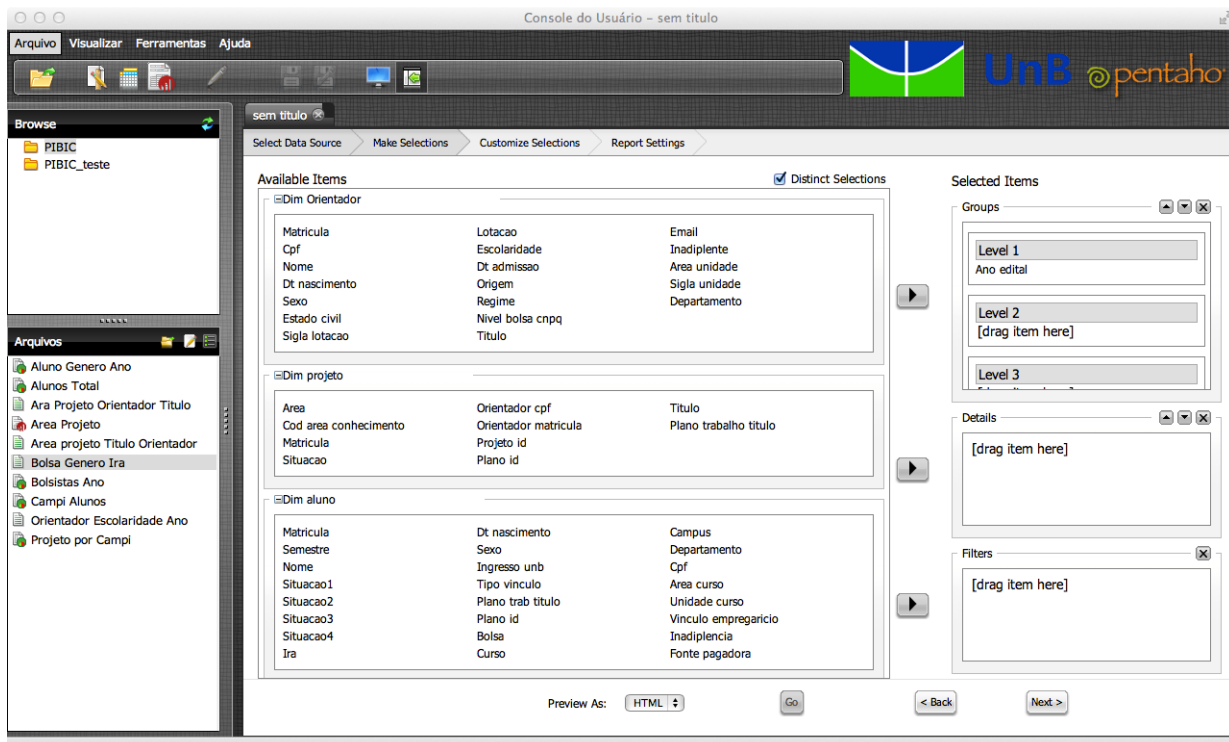


Figura 6.3: Tela de Criação de WAQR.

Fevereiro 26, 2013 @ 10:32

Area: Exatas		
Titulo	Escolaridade	Constante
DR	-	710
LD	-	4
MS	-	21
-	DOUTOR	2.483
-	GRADUADO	2
-	MESTRE	39
-	PÓS DOUTOR	7
-	-	217
Total Exatas		3.483
Area: Humanas		
Titulo	Escolaridade	Constante
DR	-	1.132
MS	-	220
-	DOUTOR	3.956
-	GRADUADO	9
-	MESTRE	134
-	NULL	12
-	PÓS DOUTOR	21
-	-	518
Total Humanas		6.002
Area: Vida		
Titulo	Escolaridade	Constante
DR	-	1.071
ES	-	12
MS	-	204
-	DOUTOR	4.383
-	GRADUADO	2

Figura 6.4: Tela de Relatório WAQR.

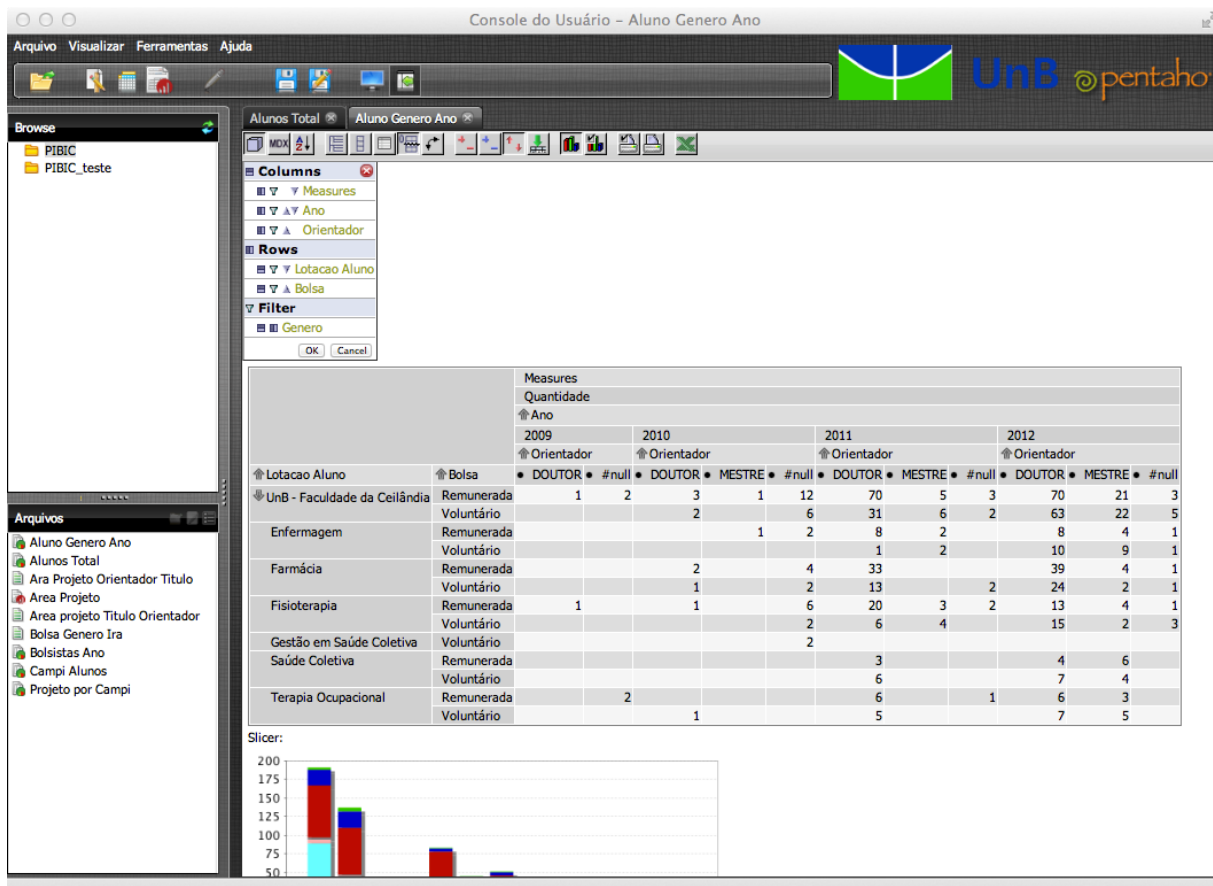


Figura 6.5: Tela de Navegação OLAP.

6.2 Qualidade dos Dados

Um problema foi detectado ao analisar o registro dos orientadores em cada ano/edital. Como pode ser visto, apesar de conter mais de 50 mil registros unindo os dois arquivos fornecidos pelo CPD, percebe-se que os registros estão duplicados e que os registros únicos são exatamente 4424. Dessa forma, verifica-se que estes registros únicos estão coerentes em relação as chaves matrícula e CPF como mostrado na Tabela 6.1 que é o resultado da *query* abaixo.

```

SELECT
COUNT(DISTINCT dim_orientador.matricula),
COUNT(DISTINCT dim_orientador.cpf),
dim_orientador.ano_edital
FROM public.dim_orientador
GROUP BY dim_orientador.ano_edital
ORDER BY dim_orientador.ano_edital

```

Tabela 6.1: Coerência das Chaves nos Registros dos Orientadores.

QTD Matrícula	Qtd CPF	Ano/Edital
0	453	2000
0	429	2001
0	433	2002
0	462	2003
383	383	2004
391	391	2005
416	416	2006
478	478	2007
450	450	2008
498	498	2009
163	163	2010
760	760	2011
885	885	2012
4424	6201	Total

6.2.1 Análise dos Registros dos Projetos e Planos de Trabalho

Projetos e planos de trabalho são registros que foram desnormalizados, pois os projetos contêm um ou mais planos. Estes registros se mostram consistentes, pois não há registros orfãos, e é constatado pela ausência de chaves nulas. Para esta análise utilizou-se a *query* indicada abaixo, que não retorna nenhum registro provando a ausência de chaves nulas.

```
SELECT projeto_sk
FROM dim_projeto
WHERE pltid is null or pro_id is null
```

Com uma investigação mais minuciosa na tabela de projetos descobriu-se que orientadores com o mesmo CPF contêm mais de uma matrícula cadastrada em um projeto. Um exemplo deste fato é mostrado na Tabela 6.2. Por este motivo adotou-se o CPF como chave estrangeira dos registros dos orientadores na tabela de projetos.

Tabela 6.2: Matrículas e CPFs de Orientadores.

Matricula	CPF	Projeto ID	Ano Edital
63	80488285100	4070	2009
1011201	80488285100	4070	2009
1010093	42185459520	1318	2005
83	42185459520	1318	2005

Os registros de matrículas distintas das tabelas *dim_aluno* e *dim_projeto*, entidades do armazém de dados são confrontados. Como um aluno pode ter preenchido mais de um

plano de trabalho, foram consultadas as matrículas distintas nas duas tabelas e percebe-se que existem muitos planos de trabalhos com matrículas de alunos que não possuem correspondentes na tabela de alunos, ou seja, faltam cadastros de alunos que fizeram planos de trabalhos. E esta falta de cadastros é mais acentuada nos anos de 2004, 2005, 2006, 2008 e 2009. A Tabela 6.3 é obtida a partir da concatenação dos resultados das seguintes *queries*:

```
SELECT COUNT (distinct matricula) as alunos, ano_ edital
FROM dim_aluno
GROUP BY ano_ edital
ORDER BY ano_ edital
```

```
SELECT COUNT (distinct matricula) as alunos_plano, ano_ edital
FROM dim_projeto
GROUP BY ano_ edital
ORDER BY ano_ edital
```

Tabela 6.3: Quantidade de Planos de Trabalho e Quantidade de Alunos.

Alunos por Plano	ano	Alunos
842	2000	842
766	2001	766
837	2002	834
921	2003	920
795	2004	77
858	2005	19
892	2006	96
1036	2007	836
1085	2008	9
1207	2009	31
1398	2010	721
1959	2011	1593
2282	2012	2121

6.3 Relatórios Gerenciais

Vale salientar que apesar de serem apresentados os seguintes relatórios, é possível ao gestor gerar relatórios *Ad-Hoc* e OLAP de acordo com a sua necessidade de informação no PUC.

6.3.1 Perfil dos Alunos

No perfil dos alunos observa-se um equilíbrio entre os sexos masculino e feminino, com uma tendência percentual um pouco maior para o sexo feminino, na maioria dos anos. Os anos de 2004 e 2008 apresentam maior tendência ao sexo masculino, porém os dados estão distorcidos por causa do espaço amostral reduzido. Os anos 2004, 2005, 2006, 2008 e 2009 apresentam distorções no resultado devido a grande quantidade de valores nulos presentes. A Figura 6.6 e a Tabela 6.4 mostram a distribuição dos alunos do sexo masculino e feminino ao longo dos anos.



Slicer:

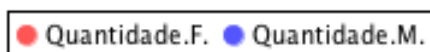


Figura 6.6: Gênero por Período.

Tabela 6.4: Gênero por Período.

Ano	F	M	#null
2000	457	407	5
2001	429	355	1
2002	450	402	3
2003	508	431	1
2004	36	43	772
2005	13	8	922
2006	75	38	911
2007	501	425	211
2008	3	7	1153
2009	25	11	1247
2010	471	311	720
2011	1028	735	374
2012	1346	957	163

6.3.2 Perfil dos Orientadores

Este relatório gerado por consulta WAQR mostra a quantidade de orientadores por áreas (vida, exatas e humanas). Também é destacado a quantidade de orientadores por escolaridade, em cada área. As siglas ES (Ensino Superior), MS (Mestre) e DR (Doutor) são referentes aos dados do Sibol. Os campos GRADUADO, MESTRE, DOUTOR e PÓS-DOUTOR são referentes aos dados do SibolWeb.

Area: Exatas

Titulo	Escolaridade	Constante
DR	-	710
LD	-	4
MS	-	21
-	DOUTOR	2483
-	GRADUADO	2
-	MESTRE	39
-	PÓS DOUTOR	7
-	-	217
Total Exatas		3483

Area: Humanas

Titulo	Escolaridade	Constante
DR	-	1132
MS	-	220
-	DOUTOR	3956
-	GRADUADO	9
-	MESTRE	134
-	NULL	12
-	PÓS DOUTOR	21
-	-	518

Total Humanas 6002
Area: Vida

Titulo	Escolaridade	Constante
DR	-	1071
ES	-	12
MS	-	204
-	DOUTOR	4383
-	GRADUADO	2
-	MESTRE	97
-	NULL	29
-	PÓS DOUTOR	7
-	-	587

Total Vida 6392
Area: -

Titulo	Escolaridade	Constante
-	DOUTOR	4
-	-	1

Total <null> 5

Grand Total 15882

6.3.3 Bolsistas no Período

Observa-se na Figura 6.7 e na Tabela 6.5 que a quantidade de bolsistas oscila entre os anos 2000 e 2005 e, a partir do ano de 2006, apresenta uma tendência de crescimento. Entre os anos de 2000 e 2006 existem somente bolsistas vinculados ao campus Darcy Ribeiro. A partir do ano de 2007 observa-se a participação de bolsistas vinculados ao Campus da Ceilândia. No ano de 2008 existem somente bolsistas vinculados ao Campus Darcy Ribeiro. Nos anos de 2009, 2010, 2011 e 2012 encontramos alunos vinculados aos demais Campus.

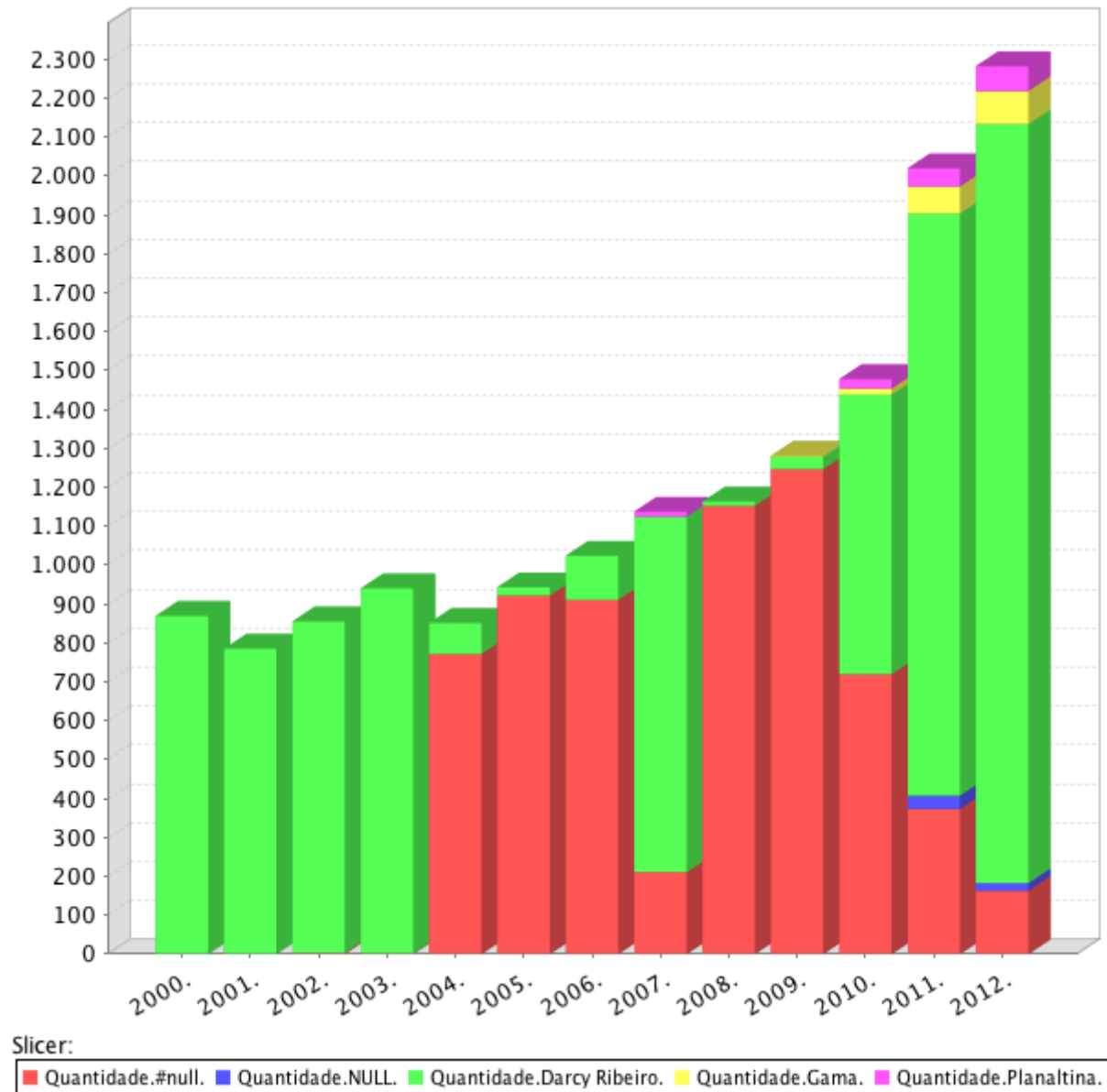


Figura 6.7: Quantidade de Discentes em Cada Campus por Período.

Tabela 6.5: Quantidade de Discentes em Cada Campus por Período.

Ano	#null	NULL	Darcy Ribeiro	Gama	Planaltina	Ceilândia
2000			869			
2001			785			
2002	3		852			
2003	1		939			
2004	772		79			
2005	922		21			
2006	911		113			
2007	211		914		12	
2008	1153		10			
2009	1247		32	1		3
2010	720		719	15	24	24
2011	374	34	1496	70	46	117
2012	163	20	1951	85	63	184

6.3.4 Número de Voluntários no Período

A nomenclatura é diferente de acordo com o período. Entre o período de 2000 e 2003 as nomenclaturas para o tipo de bolsa são as siglas NV (Não Voluntários), RN (Remunerada) e VL (Voluntário). Entre 2004 e 2012 os tipos de bolsa são as palavras Remunerada e Voluntario. Assim observa-se na Tabela 6.6 e na Figura 6.8 que em todos os anos, exceto o ano de 2006, o número de bolsas voluntárias é inferior ao número de bolsas remuneradas.

Os relatórios são contados pelos planos de trabalhos, ou seja, um aluno e um orientador podem preencher mais de um plano de trabalho, sendo apenas um por ano. Como os dados dos cadastros vieram de um sistema OLTP, foram armazenadas, apenas, a última atualização dos cadastros. Quando estes dados são referidos em anos anteriores, eles são reaproveitados. Tal caso não invalida dados do tipo sexo, que é imutável, mas caso ocorra mudança de curso ou departamento de um edital para outro, não existe este histórico sendo computado o último disponível na base.

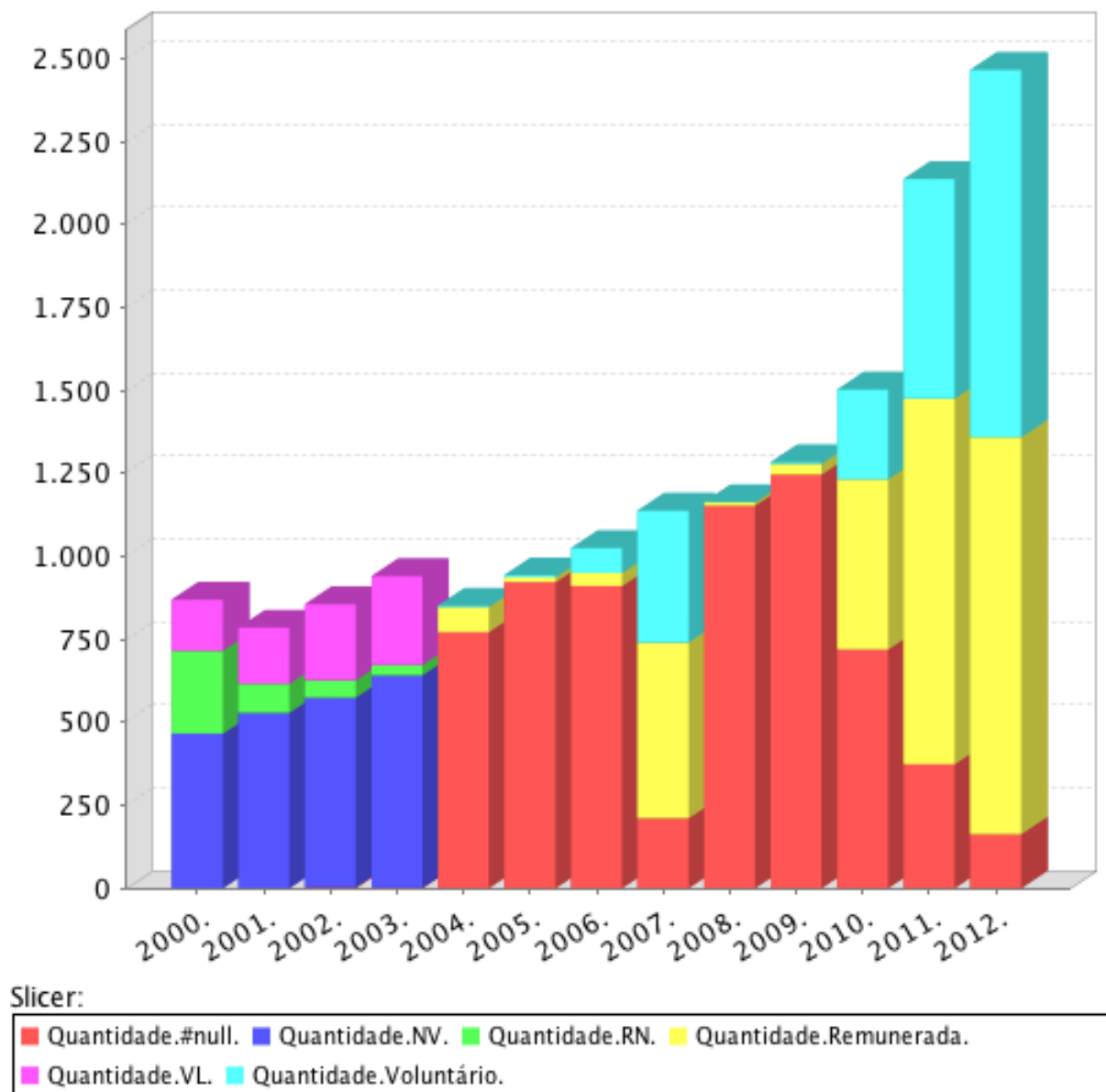


Figura 6.8: Tipo de Bolsa por Período.

Tabela 6.6: Tipo de Bolsa por Período.

Ano	#null	NV	RN	Remunerada	VL	Voluntário
2000		466	250		153	
2001		529	87		169	
2002	3	572	52		228	
2003	1	641	31		267	
2004	772			76		3
2005	922			15		6
2006	911			39		74
2007	211			530		396
2008	1153			9		1
2009	1247			32		4
2010	720			513		269
2011	374			1102		661
2012	163			1197		1106

Capítulo 7

Conclusão

O PIBIC tem entre suas metas despertar a vocação científica por incentivo aos estudantes de graduação que orientados por pesquisadores qualificados aprendem técnicas e métodos de pesquisa científica. Esse importante programa para a universidade tem guardado dados sobre seus projetos, alunos e orientadores desde do ano 2000 em sistemas com persistência em banco de dados, porém estes sistemas tem o foco na gestão rotineira dos processos, não sendo propícios a fornecer informações para auxiliar os seus gestores em tomadas de decisões e definição de políticas.

A criação do sistema de informações para o projeto de levantamento histórico do ProIC fundamentou-se nos conceitos de modelagem dimensional de banco de dados, DW e OLAP, pois estes tem o foco no armazenamento histórico e otimização na consulta dos dados agregados diferente dos sistemas transacionais que implementam a normalização dos dados para otimização de transações mantendo sempre as informações correntes.

No apoio do desenvolvimento desta solução de inteligência de negócio utilizou-se a plataforma Pentaho, que auxiliou em todas as etapas da implementação deste projeto. Desde o processo de ETL como o software PDI, a modelagem de metamodelos com PME, cubos OLAP com o PSW até a interface de geração e disponibilização dos relatórios multidimensionais no PUC. Para o armazenamento do DW foi utilizado o SGBD PostgreSQL. A plataforma Pentaho e o PostgreSQL são softwares livres.

Ao analisar o resultado do tratamento dos bancos de dados do Projeto de Iniciação Científica da UnB, é possível concluir que agregando seis bases de dados diferentes, quatro referentes ao Sibol e duas referentes o SibolWeb, cada uma com sua própria modelagem e estrutura de dados, assim como sua integração em uma modelagem dimensional, foi possível gerar informações gerenciais. Apesar da falta de cadastros dos alunos e orientadores esta metodologia nos possibilita inúmeros relacionamentos, filtros e apresentações dos dados, satisfazendo as necessidades de informações dos gestores e gerando outras num processo investigativo cíclico.

Os relatórios tem seus resultados pautados no processo que envolve os atores do ProIC, ou seja, nos planos de trabalho elaborados para os projetos de iniciação científica que totalizam 15882 registros entre os anos de 2000 a 2012. Como os cadastros fornecidos pelo CPD estão incompletos os relatórios trazem o valor *#null* que significa a quantidade de registros sem informação detalhada naquela seleção de dados. Porém, mesmo assim foi possível mostrar a evolução dos projetos de iniciação científica e agregar informação que, posteriormente, podem ser analisadas por meio de ferramentas estatísticas através da

amostragem descobrindo as margens de erro, as médias, as porcentagens sobre a população universitária e as tendências como previsto inicialmente no projeto.

Os próximos passos deste trabalho serão a instalação e configuração do PUC em um servidor na UnB, a disponibilização das rotinas de ETL para os próximos anos e a transferência de expertise para os funcionários e gestores do DPP no uso do PUC para geração e disponibilização de novos relatórios.

Algumas questões da geração de uma solução de inteligência de negócio como o controle da qualidade dos dados, modelagem multidimensional e geração de relatórios ou cubos OLAPs são tarefas que incluem arte e ciência, exigindo um conhecimento do negócio, uma prospecção de demandas e uma percepção de possibilidades de informações que os dados podem fornecer.

Nesse cenário, foi possível verificar neste trabalho que uma solução de BI com DW não é somente tecnológica, mas todo um processo que vem anterior a ele que tem que ser feito com muita responsabilidade para garantir a integridade e a confiança de seus produtos fornecendo, assim as valiosas informações que são a alma de uma inteligência de negócio.

Referências

- [1] CNPq. Página Oficial do CNPq. <http://www.cnpq.br/web/guest/pibic>. Acesso em 15 de fevereiro de 2013. 4
- [2] CRAMER, Renato. Estudo Analítico de Ferramentas Open Source para Ambientes OLAP. Universidade do Extremo Sul Catarinense, 2006. vii, viii, 13, 14, 15, 18, 19, 20
- [3] CARVALHO, Denise Birche Bomtempo e ROITMAN, Isaac. Relatório de gestão - DPP (2009-2012). Universidade de Brasília, 2012. 5, 6, 7
- [4] MELLO, João Alexandre Bonin. Uma Proposta de Modelo de Dados para Suporte ao Processamento Transacional e de Data Warehouse Simultâneamente. Universidade Federal de Santa Catarina, 2002.
- [5] SOUZA, César Alexandre e SZAFIR-GOLDSTEIN, Cláudia. Tecnologia da Informação Aplicada à Gestão Empresarial: Um Modelo para a Empresa Digital. VI SEMEAD, 2003. 10
- [6] FERREIRA, Mário César e CARVALHO, Denise Bomtempo Birche. O que é Iniciação Científica? <http://www.unb.br/noticias/unbagencia/artigo.php?id=196>. Acesso em 15 de fevereiro de 2013. 6
- [7] GOMES, Marco Antônio Boialvo. Modelação de um Data Warehouse para a Direção-Geral do Tesouro e Finanças e Implementação de um Data Mart para o Processo de Gestão Patrimonial. Universidade Nova de Lisboa, 2010. 13
- [8] PostgreSQL Global Development Group. PostgreSQL 9.0.7 Documentation. <http://www.postgresql.org/docs/9.0/interactive/index.html>. Acesso em 15 de fevereiro de 2013. 36
- [9] GUBIANI, Juçara Salete; MIRANDA, Joseane Borges; BRAGA, Marcus de Melo e TODESCO, José Leomar. Inteligência de Negócios Aplicada em Instituições de Ensino Superior. XV SIMPEP, 2008.
- [10] GUPTA, Vivek R. *An Introduction to Data Warehousing. System Services Corporation, Chicago, Illinois, 1997.* 11
- [11] INMON, Willian H. Como Construir o Dara Warehouse. Editora Campus, 2ª Edição, 1997. 12, 13, 17, 18

- [12] ITALIANO, Isabel Cristina; FERREIRA, João Eduardo e TAKAI, Osvaldo Kotaro. *Introdução a Banco de Dados*. Universidade de São Paulo, 2005. 11, 16, 17
- [13] JBoss. Página Oficial. <http://www.jboss.org>. Acesso em 19 de fevereiro de 2013. 24
- [14] JSP. Tutorial. <http://www.jsptut.com>. Acesso em 19 de fevereiro de 2013. 24
- [15] KIMBALL, Ralph e CASERTA, Joe. *The Data Warehouse ETL Toolkit*. Editora Wiley, 2ª Edição, 2004. 13, 14, 17
- [16] KIMBALL, Ralph e ROSS, Margy. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Editora Wiley, 2ª Edição, 2002. 14, 16
- [17] KIMBALL, Ralph e ROSS, Margy. *The Kimball Group Reader*. Editora Wiley, 2ª Edição, 2010.
- [18] KIMBALL, Ralph; ROSS, Margy; THORNTHWAITE, Warren; MUNDY, Joy e BECKER, Bob. *The Data Warehouse Lifecycle Toolkit*. Editora Wiley, 2ª Edição, 2008.
- [19] Microsoft. Página Oficial do Office. <http://office.microsoft.com/pt-br/access>. Acesso em 07 de março de 2013. 8
- [20] Pentaho. *Infocenter. PDI User Guide*. <http://infocenter.pentaho.com/help/index.jsp>. Acesso em 19 de fevereiro de 2013. 22, 24
- [21] Pentaho. *Mondrian Documentation*. <http://mondrian.pentaho.com/documentation/mdx.php>. Acesso em 19 de fevereiro de 2013. 29
- [22] Pentaho. *Wiki Schema Workbench*. <http://wiki.pentaho.com/display/ServerDoc1x/Publishing+an+Analysis+Schema+Using+Schema+Workbench>. Acesso em 23 de fevereiro de 2013. 21
- [23] Pentaho. *Wiki Pentaho Metadata Editor*. <http://wiki.pentaho.com/display/pmopen/Pentaho+Metadata+editor>. Acesso em 19 de fevereiro de 2013. 30
- [24] Pentaho. *Wiki Report Designer*. <http://wiki.pentaho.com/display/Reporting/Report+Designer>. Acesso em 19 de fevereiro de 2013. 22
- [25] Pentaho. *Wiki Weka*. <http://wiki.pentaho.com/display/PMOPEN/WEKA>. Acesso em 19 de fevereiro de 2013. 22
- [26] POE, Vidette; KLAUER, Patricia e BROBST, Stephen. *Building a Data Warehouse for Decision Support*. Editora Prentice Hall, 2ª Edição, 1998. 11
- [27] RAMEZ, Elmasri e NAVATHE, Shamkant B. *Sistemas de Banco de Dados*. Editora Pearson, 4ª Edição, 2005. 11
- [28] SCHEPS, Swain. *Business Intelligence For Dummies*. Editora Willey, eBook, 2008. 10

- [29] Spring Security. Página Oficial da *Spring Source*. <http://static.springsource.org/spring-security/site>. Acesso em 07 de março de 2013. 29
- [30] SHIGUNOV, Felipe. Uma Aplicação OLAP sobre a Web para Análise dos Dados do Vestibular da UFSC e Diretrizes para a sua Integração com GIS. Universidade de Santa Catarina, 2007. 18
- [31] Composite Software. *The Big Data Advantage: Take Big Advantage Of Your Data*. <http://www.compositesw.com/solutions/data-warehouse-prototyping>. Acesso em 19 de fevereiro de 2013. vii, 12
- [32] Apache Tomcat. Página Oficial. <http://tomcat.apache.org>. Acesso em 19 de fevereiro de 2013. 24
- [33] TORRES, Jorge Osvaldo De Lima. Proposta de uma Solução de Business Intelligence para a Gestão da Produção de Serviços de Tecnologias da Informação com Base no Padrão de Gerência de Redes ISO/IEC 7498-4. Universidade de Brasília, 2010. viii, 1, 22
- [34] Universidade de Brasília. Página do Decanato de Pesquisa e Pós-Graduação (DPP) da UnB. <http://www.unb.br/administracao/decanatos/dpp>. Acesso em 19 de fevereiro de 2013. 6, 40
- [35] Universidade de Brasília. Marca. <http://www.marca.unb.br>. Acesso em 23 de fevereiro de 2013. 40

Apêndice A

Relatório WAQR Titulo do Orientador por Área e Ano

Ano edital: 2000

Area: Exatas

Escolaridade	Titulo	Constante
-	DR	187
-	LD	1
-	MS	4

Total Exatas 192

Area: Humanas

Escolaridade	Titulo	Constante
-	DR	293
-	MS	61

Total Humanas 354

Area: Vida

Escolaridade	Titulo	Constante
-	DR	266
-	MS	47

Total Vida 313

Total 2000 859

Ano edital: 2001

Area: Exatas

Escolaridade	Titulo	Constante
-	DR	146
-	LD	1
-	MS	5

Total Exatas 152

Area: Humanas

Escolaridade	Titulo	Constante
-	DR	249
-	MS	59

Total Humanas 308

Area: Vida

Escolaridade	Titulo	Constante
-	DR	242
-	ES	3
-	MS	65

Total Vida 310

Total 2001 770

Ano edital: 2002

Area: Exatas

Escolaridade	Titulo	Constante
-	DR	179
-	MS	4

Total Exatas 183

Area: Humanas

Escolaridade	Titulo	Constante
-	DR	271
-	MS	52

Total Humanas 323

Area: Vida

Escolaridade	Titulo	Constante
-	DR	288
-	ES	4
-	MS	39

Total Vida 331

Total 2002 837

Ano edital: 2003

Area: Exatas

Escolaridade	Titulo	Constante
-	DR	198
-	LD	2
-	MS	8

Total Exatas 208

Area: Humanas

Escolaridade	Titulo	Constante
-	DR	319
-	MS	48
-	-	2

Total Humanas 369

Area: Vida

Escolaridade	Titulo	Constante
-	DR	275
-	ES	5
-	MS	53

Total Vida 333

Total 2003 910

Ano edital: 2004

Area: Exatas

Escolaridade	Titulo	Constante
DOUTOR	-	221
-	-	8

Total Exatas 229

Area: Humanas

Escolaridade	Titulo	Constante
DOUTOR	-	320
MESTRE	-	2
PÓS DOUTOR	-	2
-	-	15

Total Humanas 339

Area: Vida

Escolaridade	Titulo	Constante
DOUTOR	-	271
MESTRE	-	3
PÓS DOUTOR	-	1
-	-	8

Total Vida 283

Total 2004 851

Ano edital: 2005

Area: Exatas

Escolaridade	Titulo	Constante
DOUTOR	-	246
PÓS DOUTOR	-	3

Total Exatas 249

Area: Humanas

Escolaridade	Titulo	Constante
DOUTOR	-	356
GRADUADO	-	1
MESTRE	-	6
PÓS DOUTOR	-	1
-	-	11

Total Humanas 375

Area: Vida

Escolaridade	Titulo	Constante
DOUTOR	-	310
MESTRE	-	1
-	-	8

Total Vida 319

Total 2005 943

Ano edital: 2006

Area: Exatas

Escolaridade	Titulo	Constante
DOUTOR	-	204
-	-	5

Total Exatas 209

Area: Humanas

Escolaridade	Titulo	Constante
DOUTOR	-	400
MESTRE	-	1
-	-	2

Total Humanas 403

Area: Vida

Escolaridade	Titulo	Constante
DOUTOR	-	406
MESTRE	-	2
PÓS DOUTOR	-	2
-	-	2

Total Vida 412

Total 2006 1024

Ano edital: 2007

Area: Exatas

Escolaridade	Titulo	Constante
DOUTOR	-	233
GRADUADO	-	2
PÓS DOUTOR	-	1
-	-	2

Total Exatas 238

Area: Humanas

Escolaridade	Titulo	Constante
DOUTOR	-	392
GRADUADO	-	1
MESTRE	-	6
PÓS DOUTOR	-	4
-	-	2

Total Humanas 405

Area: Vida

Escolaridade	Titulo	Constante
DOUTOR	-	492
MESTRE	-	1
-	-	1

Total Vida 494

Total 2007 1137

Ano edital: 2008

Area: Exatas

Escolaridade	Titulo	Constante
--------------	--------	-----------

DOUTOR	-	213
--------	---	-----

Total Exatas		213
--------------	--	-----

Area: Humanas

Escolaridade	Titulo	Constante
--------------	--------	-----------

DOUTOR	-	433
--------	---	-----

GRADUADO	-	1
----------	---	---

MESTRE	-	1
--------	---	---

PÓS DOUTOR	-	5
------------	---	---

-	-	1
---	---	---

Total Humanas		441
---------------	--	-----

Area: Vida

Escolaridade	Titulo	Constante
--------------	--------	-----------

DOUTOR	-	502
--------	---	-----

MESTRE	-	7
--------	---	---

Total Vida		509
------------	--	-----

Total 2008		1163
------------	--	------

Ano edital: 2009

Area: Exatas

Escolaridade	Titulo	Constante
DOUTOR	-	240
PÓS DOUTOR	-	1

Total Exatas 241

Area: Humanas

Escolaridade	Titulo	Constante
DOUTOR	-	423
MESTRE	-	3
PÓS DOUTOR	-	4
-	-	21

Total Humanas 451

Area: Vida

Escolaridade	Titulo	Constante
DOUTOR	-	577
-	-	14

Total Vida 591

Total 2009 1283

Ano edital: 2010

Area: Exatas

Escolaridade	Titulo	Constante
--------------	--------	-----------

DOUTOR	-	94
--------	---	----

-	-	187
---	---	-----

Total Exatas		281
--------------	--	-----

Area: Humanas

Escolaridade	Titulo	Constante
--------------	--------	-----------

DOUTOR	-	98
--------	---	----

-	-	393
---	---	-----

Total Humanas		491
---------------	--	-----

Area: Vida

Escolaridade	Titulo	Constante
--------------	--------	-----------

DOUTOR	-	225
--------	---	-----

MESTRE	-	1
--------	---	---

-	-	499
---	---	-----

Total Vida		725
------------	--	-----

Area: -

Escolaridade	Titulo	Constante
--------------	--------	-----------

DOUTOR	-	4
--------	---	---

-	-	1
---	---	---

Total <null>		5
--------------	--	---

Total 2010		1502
------------	--	------

Ano edital: 2011

Area: Exatas

Escolaridade	Titulo	Constante
DOUTOR	-	464
MESTRE	-	22
PÓS DOUTOR	-	2
-	-	8

Total Exatas 496

Area: Humanas

Escolaridade	Titulo	Constante
DOUTOR	-	722
GRADUADO	-	2
MESTRE	-	35
NULL	-	3
-	-	32

Total Humanas 794

Area: Vida

Escolaridade	Titulo	Constante
DOUTOR	-	779
MESTRE	-	27
NULL	-	4
-	-	37

Total Vida 847

Total 2011 2137

Ano edital: 2012

Area: Exatas

Escolaridade	Titulo	Constante
DOUTOR	-	568
MESTRE	-	17
-	-	7

Total Exatas 592

Area: Humanas

Escolaridade	Titulo	Constante
DOUTOR	-	812
GRADUADO	-	4
MESTRE	-	80
NULL	-	9
PÓS DOUTOR	-	5
-	-	39

Total Humanas 949

Area: Vida

Escolaridade	Titulo	Constante
DOUTOR	-	821
GRADUADO	-	2
MESTRE	-	55
NULL	-	25
PÓS DOUTOR	-	4
-	-	18

Total Vida 925

Total 2012 2466

Grand Total 15882

Apêndice B

Relatório WAQR da Média do IRA pelo Ano

Ano edital: 2000

Bolsa	Sexo	Ira	Constante
NV	F	3.9270536111	252
NV	M	3.7363122642	212
NV	-	3.51465	2
RN	F	3.9723438462	130
RN	M	3.827885	120
VL	F	3.8704826667	75
VL	M	3.9428801333	75
VL	-	4.1887333333	3
Total 2000		3.8725426068	869

Ano edital: 2001

Bolsa	Sexo	Ira	Constante
NV	F	4.0556890813	283
NV	M	3.8700449796	245
NV	-	3.2051	1
RN	F	3.9776326087	46
RN	M	3.912297561	41
VL	F	3.977783	100
VL	M	4.052663913	69
Total 2001		3.8644587348	785

Ano edital: 2002

Bolsa	Sexo	Ira	Constante
NV	F	4.0992886039	308
NV	M	3.9297371212	264
RN	F	4.006576	25
RN	M	3.9304022222	27
VL	F	4.0560606838	117
VL	M	3.8908001802	111
-	-	-	3
Total 2002		3.9854774685	855

Ano edital: 2003

Bolsa	Sexo	Ira	Constante
NV	F	4.1527142857	350
NV	M	3.929380756	291
RN	F	4.2356578947	19
RN	M	3.7248716667	12
VL	F	4.1472798561	139
VL	M	3.8948195313	128
-	-	-	1
Total 2003		4.0141206651	940

Ano edital: 2004

Bolsa	Sexo	Ira	Constante
Remunerada	F	-	34
Remunerada	M	-	42
Voluntário	F	-	2
Voluntário	M	-	1
-	-	-	772
Total 2004			851

Ano edital: 2005

Bolsa	Sexo	Ira	Constante
Remunerada	F	-	10
Remunerada	M	-	5
Voluntário	F	-	3
Voluntário	M	-	3
-	-	-	922
Total 2005			943

Ano edital: 2006

Bolsa	Sexo	Ira	Constante
Remunerada	F	-	26
Remunerada	M	3.3686	13
Voluntário	F	3.439	49
Voluntário	M	-	25
-	-	-	911
Total 2006		3.4038	1024

Ano edital: 2007

Bolsa	Sexo	Ira	Constante
Remunerada	F	4.2014210526	286
Remunerada	M	3.8702863636	244
Voluntário	F	3.57145	215
Voluntário	M	3.81795	181
-	-	-	211
Total 2007		3.8652768541	1137

Ano edital: 2008

Bolsa	Sexo	Ira	Constante
Remunerada	F	-	2
Remunerada	M	4.0688	7
Voluntário	F	-	1
-	-	-	1153
Total 2008		4.0688	1163

Ano edital: 2009

Bolsa	Sexo	Ira	Constante
Remunerada	F	3.830225	24
Remunerada	M	3.9296666667	8
Voluntário	F	-	1
Voluntário	M	4.224	3
-	-	-	1247
Total 2009		3.9946305556	1283

Ano edital: 2010

Bolsa	Sexo	Ira	Constante
Remunerada	F	3.9927079832	306
Remunerada	M	3.7921287356	207
Voluntário	F	4.0735572519	165
Voluntário	M	3.9442337349	104
-	-	-	720
Total 2010		3.9506569264	1502

Ano edital: 2011

Bolsa	Sexo	Ira	Constante
Remunerada	F	4.0076850863	650
Remunerada	M	3.8084323596	452
Voluntário	F	4.0296532432	378
Voluntário	M	3.87026787	283
-	-	-	374
Total 2011		3.9290096398	2137

Ano edital: 2012

Bolsa	Sexo	Ira	Constante
Remunerada	F	3.9246502865	709
Remunerada	M	3.7581993827	488
Voluntário	F	3.965048265	637
Voluntário	M	3.7985002151	469
-	-	-	163
Total 2012		3.8615995373	2466

Grand Total		3.9044000877	15955
-------------	--	--------------	-------

Apêndice C

Mapa Lógico de Dados

Tabela C.1: Gestão

Sistema de Origem	Prioridade
SIBOL	
SIBOL	
SIBOL	
SIBOL	
SIBOL	
SIBOL	
SIBOL	
SIBOL	
SIBOL	
SIBOL	
SIBOLWEB	
SIBOLWEB	
SIBOLWEB	
SIBOLWEB	
SIBOLWEB	
SIBOLWEB	
SIBOLWEB	
SIBOLWEB	
SIBOLWEB	
SIBOLWEB	
SIBOLWEB	

Tabela C.2: Indentificação do DW

DFS	Tabelas	Coluna	Tipo	SCD
	DIM_ORIENTADOR	CELULAR	integer	
	DIM_ORIENTADOR	CEP	integer	
	DIM_ORIENTADOR	CIDADEEENDEREÇO	string	
	DIM_ORIENTADOR	CPF	string	
	DIM_ORIENTADOR	DTNASCIMENTO	date	
	DIM_ORIENTADOR	ENDERECO	string	
	DIM_ORIENTADOR	ESCOLARIDADE	string	
	DIM_ORIENTADOR	FONES	integer	
	DIM_ORIENTADOR	LOTACAO	string	
	DIM_ORIENTADOR	NOME	string	
	DIM_ALUNO	CPF	string	
	DIM_ALUNO	DATA_NASCIMENTO	data	
	DIM_ALUNO	EMAIL	STRING	
	DIM_ALUNO	INGRESSO_UNB	string	
	DIM_ALUNO	IRA	float	
	DIM_ALUNO	MATRICULA	string	
	DIM_ALUNO	NOME	STRING	
	DIM_ALUNO	SEXO	string	
	DIM_ALUNO	VINCULO_EMPREGO	boolean	
	DIM_ALUNO	BOLSA	string	
	DIM_ALUNO	CPF	string	
	DIM_ALUNO	DATA_NASCIMENTO	data	
	DIM_ALUNO	EMAIL	STRING	
	DIM_ALUNO	INGRESSO_UNB	string	
	DIM_ALUNO	IRA	float	
	DIM_ALUNO	MATRICULA	string	
	DIM_ALUNO	NOME	string	
	DIM_ALUNO	SEXO	string	
	DIM_ALUNO	VINCULO_EMPREGO	boolean	
	DIM_ALUNO	BOLSA	string	
	DIM_ALUNO	CPF	string	
	DIM_ALUNO	DATA_NASCIMENTO	data	
	DIM_ALUNO	EMAIL	STRING	
	DIM_ALUNO	INGRESSO_UNB	string	
	DIM_ALUNO	IRA	float	
	dim_aluno	MATRICULA	string	
	DIM_ALUNO	NOME	string	
	DIM_ALUNO	SEXO	string	
	DIM_ALUNO	VINCULO_EMPREGO	boolean	
	DIM_ALUNO	BOLSA	string	
	DIM_ALUNO	CPF	string	
	DIM_ALUNO	DATA_NASCIMENTO	data	

	DIM_ALUNO	EMAIL	STRING	
	DIM_ALUNO	INGRESSO_UNB	string	
	DIM_ALUNO	IRA	float	
	DIM_ALUNO	MATRICULA	string	
	DIM_ALUNO	NOME	string	
	DIM_ALUNO	SEXO	char	
	DIM_ALUNO	VINCULO_EMPREGO	boolean	
	DIM_ORIENTADOR	ANOEDITAL	integer	
	DIM_ORIENTADOR	CELULAR	integer	
	DIM_ORIENTADOR	CEP	integer	
	DIM_ORIENTADOR	CIDADEENDERECO	string	
	DIM_ORIENTADOR	CPF	integer	
	DIM_ORIENTADOR	DTADMISSAO	date	
	DIM_ORIENTADOR	DTNASCIMENTAO	date	
	DIM_ORIENTADOR	ENDERECO	string	
	DIM_ORIENTADOR	ESCOLARIDADE	string	
	DIM_ORIENTADOR	ESTADOCIVIL	string	
	DIM_ORIENTADOR	FONES	integer	
	DIM_ORIENTADOR	LOTACAO	string	
	DIM_ORIENTADOR	MATRICULA	integer	
	DIM_ORIENTADOR	NIVELBOLSACNPQ	string	
	DIM_ORIENTADOR	NOME	string	
	DIM_ORIENTADOR	ORIGEM	string	
	DIM_ORIENTADOR	PRODUCAOCIENTIFICA	string	
	DIM_ORIENTADOR	QUANTIDADE	integer	
	DIM_ORIENTADOR	REGIME	string	
	DIM_ORIENTADOR	SEXO	char	
	DIM_ORIENTADOR	SIGLALOTACAO	string	
	DIM_ALUNO	BOLSA	string	
	DIM_ALUNO	CAMPUS	string	
	DIM_ALUNO	CURSO	string	
	DIM_ALUNO	DATA_NASCIMENTO	date	
	DIM_ALUNO	DEPARTAMENTO	string	
	DIM_ALUNO	EDITAL_ANO	integer	
	DIM_ALUNO	INGRESSO_UNB	string	
	DIM_ALUNO	IRA	float	
	DIM_ALUNO	MATRICULA	string	
	DIM_ALUNO	NOME	string	
	DIM_ALUNO	PLANO_TRABALHO_ID	integer	
	DIM_ALUNO	SEXO	String	
	DIM_ALUNO	SITUACAO1	string	
	DIM_ALUNO	SITUACAO2	string	
	DIM_ALUNO	SITUACAO3	string	
	DIM_ALUNO	SITUACAO4	string	
	DIM_ALUNO	VINCULO_UNB	string	

SIC_PIBIC_2000	Alunos	RgNumeroAluno	string
SIC_PIBIC_2000	Alunos	RgOrgaoEmissorAluno	string
SIC_PIBIC_2000	Alunos	RgUfAluno	string
SIC_PIBIC_2000	Alunos	Endereco	string
SIC_PIBIC_2000	Alunos	Bairro	string
SIC_PIBIC_2000	Alunos	Cidade	string
SIC_PIBIC_2000	Alunos	Uf	string
SIC_PIBIC_2000	Alunos	Cep	string
SIC_PIBIC_2000	Alunos	TelefoneAluno	string
SIC_PIBIC_2000	Alunos	TelefoneCelularAluno	string
SIC_PIBIC_2000	Alunos	RamalAluno	string
SIC_PIBIC_2000	Alunos	AgenciaAluno	string
SIC_PIBIC_2000	Alunos	ContaAluno	string
SIC_PIBIC_2000	Alunos	InadimplenciaPendente	boolean
SIC_PIBIC_2001	Orientadores	CpfOrientador	string
SIC_PIBIC_2001	Orientadores	NomeOrientador	string
SIC_PIBIC_2001	Orientadores	TitulacaoOrientador	string
SIC_PIBIC_2001	Orientadores	RgNumeroOrientador	string
SIC_PIBIC_2001	Orientadores	RgOrgaoEmissorOrientador	string
SIC_PIBIC_2001	Orientadores	RgUfOrientador	string
SIC_PIBIC_2001	Orientadores	CodUnidadeOrientador	integer
SIC_PIBIC_2001	Orientadores	FuncaoOrientador	string
SIC_PIBIC_2001	Orientadores	Endereco	string
SIC_PIBIC_2001	Orientadores	Bairro	string
SIC_PIBIC_2001	Orientadores	Cidade	string
SIC_PIBIC_2001	Orientadores	Uf	string
SIC_PIBIC_2001	Orientadores	Cep	string
SIC_PIBIC_2001	Orientadores	Telefone	string
SIC_PIBIC_2001	Orientadores	TelefoneCelular	string
SIC_PIBIC_2001	Orientadores	Ramal	string
SIC_PIBIC_2001	Orientadores	Email	string
SIC_PIBIC_2001	Orientadores	SubAreaCnpqOrientador	string
SIC_PIBIC_2001	Orientadores	Ativo	boolean
SIC_PIBIC_2001	Orientadores	DataInativo	date
SIC_PIBIC_2001	Orientadores	OrientadorInadimplente	boolean
SIC_PIBIC_2001	Alunos	BolsaRemunerada	boolean
SIC_PIBIC_2001	Alunos	CpfAluno	string
SIC_PIBIC_2001	Alunos	DataNascimento	date
SIC_PIBIC_2001	Alunos	EmailAluno	string
SIC_PIBIC_2001	Alunos	SemestreIngressoo	string
SIC_PIBIC_2001	Alunos	IraAluno	double

Tabela C.3: Fontes (Tabela e Tipo)

Database	Tabela	Coluna	Tipo
SIC_PIBIC_2000	Planos de Trabalho	NumeroPlanoTrabalho	string
SIC_PIBIC_2000	Planos de Trabalho	MatriculaAluno	string
SIC_PIBIC_2000	Planos de Trabalho	CpfOrientador	string
SIC_PIBIC_2000	Planos de Trabalho	NumeroProjetoRelacionado	string
SIC_PIBIC_2000	Planos de Trabalho	CondicaoPlanoTrabalho	string
SIC_PIBIC_2000	Planos de Trabalho	TipoBolsa	string
SIC_PIBIC_2000	Planos de Trabalho	FontePagadora	string
SIC_PIBIC_2000	Orientadores	RgNumeroOrientador	string
SIC_PIBIC_2000	Orientadores	RgOrgaoEmissorOrientador	string
SIC_PIBIC_2000	Orientadores	RgUfOrientador	string
SIC_PIBIC_2000	Orientadores	TelefoneCelular	string
SIC_PIBIC_2000	Orientadores	Cep	string
SIC_PIBIC_2000	Orientadores	Cidade	string
SIC_PIBIC_2000	Orientadores	Uf	string
SIC_PIBIC_2000	Orientadores	CpfOrientador	string
SIC_PIBIC_2000	Orientadores	Endereco	string
SIC_PIBIC_2000	Orientadores	Bairro	string
SIC_PIBIC_2000	Orientadores	TitulacaoOrientador	string
SIC_PIBIC_2000	Orientadores	Telefone	string
SIC_PIBIC_2000	Orientadores	CodUnidadeOrientador	integer
SIC_PIBIC_2000	Orientadores	NomeOrientador	string
SIC_PIBIC_2000	Orientadores	FuncaoOrientador	string
SIC_PIBIC_2000	Orientadores	Ramal	string
SIC_PIBIC_2000	Orientadores	Email	string
SIC_PIBIC_2000	Orientadores	SubAreaCnpqOrientador	string
SIC_PIBIC_2000	Orientadores	Ativo	boolean
SIC_PIBIC_2000	Orientadores	DataInativo	date
SIC_PIBIC_2000	Orientadores	OrientadorInadimplente	boolean
SIC_PIBIC_2000	Alunos	CpfAluno	string
SIC_PIBIC_2000	Alunos	DataNascimento	date
SIC_PIBIC_2000	Alunos	EmailAluno	string
SIC_PIBIC_2000	Alunos	SemestreIngressoo	string
SIC_PIBIC_2000	Alunos	IraAluno	double
SIC_PIBIC_2000	Alunos	MatriculaAluno	string
SIC_PIBIC_2000	Alunos	NomeAluno	string
SIC_PIBIC_2000	Alunos	Sexo	string
SIC_PIBIC_2000	Alunos	VinculoEmpregaticio	boolean
SIC_PIBIC_2000	Alunos	CodigoCurso	Integer
SIC_PIBIC_2000	Alunos	AnoIngresso	Integer
SIC_PIBIC_2000	Alunos	Reprovações Anteriores	integer
SIC_PIBIC_2000	Alunos	Reprovações OK	boolean

SIC_PIBIC_2001	Alunos	MatriculaAluno	string
SIC_PIBIC_2001	Alunos	NomeAluno	string
SIC_PIBIC_2001	Alunos	Sexo	string
SIC_PIBIC_2001	Alunos	VinculoEmpregaticio	boolean
SIC_PIBIC_2001	Alunos	CodigoCurso	Integer
SIC_PIBIC_2001	Alunos	AnoIngresso	Integer
SIC_PIBIC_2001	Alunos	ReprovoacoesAnteriores	integer
SIC_PIBIC_2001	Alunos	ReprovoacoesOK	boolean
SIC_PIBIC_2001	Alunos	RgNumeroAluno	string
SIC_PIBIC_2001	Alunos	RgOrgaoEmissorAluno	string
SIC_PIBIC_2001	Alunos	RgUfAluno	string
SIC_PIBIC_2001	Alunos	Endereco	string
SIC_PIBIC_2001	Alunos	Bairro	string
SIC_PIBIC_2001	Alunos	Cidade	string
SIC_PIBIC_2001	Alunos	Uf	string
SIC_PIBIC_2001	Alunos	Cep	string
SIC_PIBIC_2001	Alunos	TelefoneAluno	string
SIC_PIBIC_2001	Alunos	TelefoneCelularAluno	string
SIC_PIBIC_2001	Alunos	RamalAluno	string
SIC_PIBIC_2001	Alunos	AgenciaAluno	string
SIC_PIBIC_2001	Alunos	ContaAluno	string
SIC_PIBIC_2001	Alunos	InadimplenciaPendente	boolean
SIC_PIBIC_2001	Planos	NumeroPlanoTrabalho	string
SIC_PIBIC_2001	Planos	MatriculaAluno	string
SIC_PIBIC_2001	Planos	CpfOrientador	string
SIC_PIBIC_2001	Planos	NumeroProjetoRelacionado	string
SIC_PIBIC_2001	Planos	AreaPlanoTrabalho	string
SIC_PIBIC_2001	Planos	MotivoExclusao	string
SIC_PIBIC_2001	Planos	RecursoApresentado	boolean
SIC_PIBIC_2001	Planos	PedidoReconsideracao	string
SIC_PIBIC_2001	Planos	CondicaoPlanoTrabalho	string
SIC_PIBIC_2001	Planos	TipoBolsa	string
SIC_PIBIC_2001	Planos	FontePagadora	string
SIC_PIBIC_2001	Planos	Palavra_chave1	string
SIC_PIBIC_2001	Planos	Palavra_chave2	string
SIC_PIBIC_2001	Planos	Palavra_chave3	string
SIC_PIBIC_2001	Planos	RelatorioParcialEntregue	boolean
SIC_PIBIC_2001	Planos	ApresentacaoOralOk	boolean
SIC_PIBIC_2001	Planos	ResumoEntregue	boolean
SIC_PIBIC_2001	Planos	RelatorioFinalEntregue	boolean

SIC_PIBIC_2001	Planos	ApresentacaoCongresso	boolean
SIC_PIBIC_2001	Planos	NomeAssessorExtra	string
SIC_PIBIC_2001	Planos	AvaliacaoInserida1	boolean
SIC_PIBIC_2001	Planos	AvaliacaoInserida2	boolean
SIC_PIBIC_2001	Planos	Inadimplente	boolean
SIC_PIBIC_2002	Orientadores	CpfOrientador	string
SIC_PIBIC_2002	Orientadores	NomeOrientador	string
SIC_PIBIC_2002	Orientadores	TituloOrientador	string
SIC_PIBIC_2002	Orientadores	RgNumeroOrientador	string
SIC_PIBIC_2002	Orientadores	RgOrgaoEmissorOrientador	string
SIC_PIBIC_2002	Orientadores	RgUfOrientador	string
SIC_PIBIC_2002	Orientadores	CodUnidadeOrientador	integer
SIC_PIBIC_2002	Orientadores	FuncaoOrientador	string
SIC_PIBIC_2002	Orientadores	Endereco	string
SIC_PIBIC_2002	Orientadores	Bairro	string
SIC_PIBIC_2002	Orientadores	Cidade	string
SIC_PIBIC_2002	Orientadores	Uf	string
SIC_PIBIC_2002	Orientadores	Cep	string
SIC_PIBIC_2002	Orientadores	Telefone	string
SIC_PIBIC_2002	Orientadores	TelefoneCelular	string
SIC_PIBIC_2002	Orientadores	Ramal	string
SIC_PIBIC_2002	Orientadores	Email	string
SIC_PIBIC_2002	Orientadores	SubAreaCnpqOrientador	string
SIC_PIBIC_2002	Orientadores	Ativo	boolean
SIC_PIBIC_2002	Orientadores	OrientadorInadimplente	boolean
SIC_PIBIC_2002	Alunos	BolsaRemunerada	boolean
SIC_PIBIC_2002	Alunos	CpfAluno	string
SIC_PIBIC_2002	Alunos	EmailAluno	string
SIC_PIBIC_2002	Alunos	SemestreIngresso	string
SIC_PIBIC_2002	Alunos	IraAluno	double
SIC_PIBIC_2002	Alunos	MatriculaAluno	string
SIC_PIBIC_2002	Alunos	NomeAluno	string
SIC_PIBIC_2002	Alunos	Sexo	string
SIC_PIBIC_2002	Alunos	VinculoEmpregaticio	boolean
SIC_PIBIC_2002	Alunos	CodigoCurso	Integer
SIC_PIBIC_2002	Alunos	AnoIngresso	Integer
SIC_PIBIC_2002	Alunos	ReprovaçõesAnteriores	integer
SIC_PIBIC_2002	Alunos	ReprovaçõesOK	boolean
SIC_PIBIC_2002	Alunos	RgNumeroAluno	string
SIC_PIBIC_2002	Alunos	RgOrgaoEmissorAluno	string
SIC_PIBIC_2002	Alunos	RgUfAluno	string

SIC_PIBIC_2002	Alunos	Endereco	string
SIC_PIBIC_2002	Alunos	Bairro	string
SIC_PIBIC_2002	Alunos	Cidade	string
SIC_PIBIC_2002	Alunos	Uf	string
SIC_PIBIC_2002	Alunos	Cep	string
SIC_PIBIC_2002	Alunos	TelefoneAluno	string
SIC_PIBIC_2002	Alunos	TelefoneCelularAluno	string
SIC_PIBIC_2002	Alunos	RamalAluno	string
SIC_PIBIC_2002	Alunos	AgenciaAluno	string
SIC_PIBIC_2002	Alunos	ContaAluno	string
SIC_PIBIC_2002	Alunos	InadimplenciaPendente	boolean
SIC_PIBIC_2002	Planos	NumeroPlanoTrabalho	string
SIC_PIBIC_2002	Planos	MatriculaAluno	string
SIC_PIBIC_2002	Planos	CpfOrientador	string
SIC_PIBIC_2002	Planos	NumeroProjetoRelacionado	string
SIC_PIBIC_2002	Planos	TituloPlanoTrabalho	string
SIC_PIBIC_2002	Planos	AreaPlanoTrabalho	string
SIC_PIBIC_2002	Planos	MotivoExclusao	string
SIC_PIBIC_2002	Planos	RecursoApresentado	boolean
SIC_PIBIC_2002	Planos	CondicaoPlanoTrabalho	string
SIC_PIBIC_2002	Planos	TipoBolsa	string
SIC_PIBIC_2002	Planos	FontePagadora	string
SIC_PIBIC_2002	Planos	Palavra_chave1	string
SIC_PIBIC_2002	Planos	palavra_chave2	string
SIC_PIBIC_2002	Planos	palavra_chave3	string
SIC_PIBIC_2002	Planos	RelatorioParcialEntregue	boolean
SIC_PIBIC_2002	Planos	ApresentacaoOralOk	boolean
SIC_PIBIC_2002	Planos	ResumoEntregue	boolean
SIC_PIBIC_2002	Planos	RelatorioFinalEntregue	boolean
SIC_PIBIC_2002	Planos	ApresentacaoCongresso	boolean
SIC_PIBIC_2002	Planos	NomeAssessorExtra	string
SIC_PIBIC_2002	Planos	AvaliacaoInserida1	boolean
SIC_PIBIC_2002	Planos	AvaliacaoInserida2	boolean
SIC_PIBIC_2002	Planos	Inadimplente	boolean
SIC_PIBIC_2003	Planos de Trabalho	NumeroPlanoTrabalho	string
SIC_PIBIC_2003	Planos de Trabalho	MatriculaAluno	string
SIC_PIBIC_2003	Planos de Trabalho	CpfOrientador	string
SIC_PIBIC_2003	Planos de Trabalho	NumeroProjetoRelacionado	string
SIC_PIBIC_2003	Planos de Trabalho	CondicaoPlanoTrabalho	string
SIC_PIBIC_2003	Planos de Trabalho	TipoBolsa	string
SIC_PIBIC_2003	Planos de Trabalho	FontePagadora	string

SIC_PIBIC_2003	Orientadores	CpfOrientador	string
SIC_PIBIC_2003	Orientadores	NomeOrientador	string
SIC_PIBIC_2003	Orientadores	TitulacaoOrientador	string
SIC_PIBIC_2003	Orientadores	RgNumeroOrientador	string
SIC_PIBIC_2003	Orientadores	RgOrgaoEmissorOrientador	string
SIC_PIBIC_2003	Orientadores	RgUfOrientador	string
SIC_PIBIC_2003	Orientadores	CodUnidadeOrientador	integer
SIC_PIBIC_2003	Orientadores	FuncaoOrientador	string
SIC_PIBIC_2003	Orientadores	Endereco	string
SIC_PIBIC_2003	Orientadores	Bairro	string
SIC_PIBIC_2003	Orientadores	Cidade	string
SIC_PIBIC_2003	Orientadores	Uf	string
SIC_PIBIC_2003	Orientadores	Cep	string
SIC_PIBIC_2003	Orientadores	Telefone	string
SIC_PIBIC_2003	Orientadores	TelefoneCelular	string
SIC_PIBIC_2003	Orientadores	Ramal	string
SIC_PIBIC_2003	Orientadores	Email	string
SIC_PIBIC_2003	Orientadores	SubAreaCnpqOrientador	string
SIC_PIBIC_2003	Orientadores	Ativo	boolean
SIC_PIBIC_2003	Orientadores	OrientadorInadimplente	boolean
SIC_PIBIC_2003	Alunos	BolsaRemunerada	booleana
SIC_PIBIC_2003	Alunos	CpfAluno	string
SIC_PIBIC_2003	Alunos	DataNascimento	date
SIC_PIBIC_2003	Alunos	EmailAluno	string
SIC_PIBIC_2003	Alunos	SemestreIngressoo	string
SIC_PIBIC_2003	Alunos	IraAluno	double
SIC_PIBIC_2003	Alunos	MatriculaAluno	string
SIC_PIBIC_2003	Alunos	NomeAluno	string
SIC_PIBIC_2003	Alunos	Sexo	string
SIC_PIBIC_2003	Alunos	VinculoEmpregaticio	boolean
SIC_PIBIC_2003	Alunos	CodigoCurso	Integer
SIC_PIBIC_2003	Alunos	AnoIngresso	Integer
SIC_PIBIC_2003	Alunos	ReprovaçõesAnteriores	integer
SIC_PIBIC_2003	Alunos	ReprovaçõesOK	boolean
SIC_PIBIC_2003	Alunos	RgNumeroAluno	string
SIC_PIBIC_2003	Alunos	RgOrgaoEmissorAluno	string
SIC_PIBIC_2003	Alunos	RgUfAluno	string
SIC_PIBIC_2003	Alunos	Endereco	string
SIC_PIBIC_2003	Alunos	Bairro	string
SIC_PIBIC_2003	Alunos	Cidade	string
SIC_PIBIC_2003	Alunos	Uf	string

SIC_PIBIC_2003	Alunos	Cep	string
SIC_PIBIC_2003	Alunos	TelefoneAluno	string
SIC_PIBIC_2003	Alunos	TelefoneCelularAluno	string
SIC_PIBIC_2003	Alunos	RamalAluno	string
SIC_PIBIC_2003	Alunos	AgenciaAluno	string
SIC_PIBIC_2003	Alunos	ContaAluno	string
SIC_PIBIC_2003	Alunos	InadimplenciaPendente	boolean
SIC_PIBIC_2005-2012	PlanoDeTrabalho	ProId	integer
SIC_PIBIC_2005-2012	PlanoDeTrabalho	PltId	integer
SIC_PIBIC_2005-2012	PlanoDeTrabalho	Matricula	string
SIC_PIBIC_2005-2012	PlanoDeTrabalho	Situacao	string
SIC_PIBIC_2005-2012	Docentes	EdiAnoEdital	integer
SIC_PIBIC_2005-2012	Docentes	oricelular	integer
SIC_PIBIC_2005-2012	Docentes	oricep	integer
SIC_PIBIC_2005-2012	Docentes	oricidadeendereço	string
SIC_PIBIC_2005-2012	Docentes	oricpf	integer
SIC_PIBIC_2005-2012	Docentes	escolaridade	string
SIC_PIBIC_2005-2012	Docentes	oriestadocivil	string
SIC_PIBIC_2005-2012	Docentes	orifoneres	integer
SIC_PIBIC_2005-2012	Docentes	orimatricula	integer
SIC_PIBIC_2005-2012	Docentes	opnivelbolsaprodcnpq	string
SIC_PIBIC_2005-2012	Docentes	oriorigem	string
SIC_PIBIC_2005-2012	Docentes	Quantidade	integer
SIC_PIBIC_2005-2012	Docentes	regime	string
SIC_PIBIC_2005-2012	Docentes	orisexo	char
SIC_PIBIC_2005-2012	Docentes	siglalotacao	string
SIC_PIBIC_2005-2012	Discentes	TipoBolsa	string
SIC_PIBIC_2005-2012	Discentes	Campus	string
SIC_PIBIC_2005-2012	Discentes	EdiAnoEdital	integer
SIC_PIBIC_2005-2012	Discentes	aluperingunb	integer
SIC_PIBIC_2005-2012	Discentes	AluIRA	float
SIC_PIBIC_2005-2012	Discentes	AluMatricula	string
SIC_PIBIC_2005-2012	Discentes	PltPrpId	integer
SIC_PIBIC_2005-2012	Discentes	alusexo	char
SIC_PIBIC_2005-2012	Discentes	Situação1	string
SIC_PIBIC_2005-2012	Discentes	Situação2	string
SIC_PIBIC_2005-2012	Discentes	Situação3	string
SIC_PIBIC_2005-2012	Discentes	Situação4	string
SIC_PIBIC_2005-2012	Discentes	TipoVínculo	string
SIC_PIBIC_2005-2012	Discentes	Semestre	integer
SIC_PIBIC_2005-2012	Discentes	PltTitulo	string

Tabela C.4: Fontes (Tamanho e Precisão)

Database	Coluna	Tamanho	Precisão
SIC_PIBIC_2000	NumeroPlanoTrabalho	6	
SIC_PIBIC_2000	MatriculaAluno	9	
SIC_PIBIC_2000	CpfOrientador	15	
SIC_PIBIC_2000	NumeroProjetoRelacionado	6	
SIC_PIBIC_2000	CondicaoPlanoTrabalho	20	
SIC_PIBIC_2000	TipoBolsa	20	
SIC_PIBIC_2000	FontePagadora	50	
SIC_PIBIC_2000	RgNumeroOrientador	18	
SIC_PIBIC_2000	RgOrgaoEmissorOrientador	10	
SIC_PIBIC_2000	RgUfOrientador	2	
SIC_PIBIC_2000	TelefoneCelular	20	
SIC_PIBIC_2000	Cep	12	
SIC_PIBIC_2000	Cidade	50	
SIC_PIBIC_2000	Uf	3	
SIC_PIBIC_2000	CpfOrientador	15	
SIC_PIBIC_2000	Endereco	70	
SIC_PIBIC_2000	Bairro	30	
SIC_PIBIC_2000	TitulacaoOrientador	5	
SIC_PIBIC_2000	Telefone	20	
SIC_PIBIC_2000	CodUnidadeOrientador	2	
SIC_PIBIC_2000	NomeOrientador	50	
SIC_PIBIC_2000	FuncaoOrientador	5	
SIC_PIBIC_2000	Ramal	10	
SIC_PIBIC_2000	Email	50	
SIC_PIBIC_2000	SubAreaCnpqOrientador	20	
SIC_PIBIC_2000	Ativo	1	
SIC_PIBIC_2000	DataInativo		
SIC_PIBIC_2000	OrientadorInadimplente	1	
SIC_PIBIC_2000	CpfAluno	15	
SIC_PIBIC_2000	DataNascimento		
SIC_PIBIC_2000	EmailAluno	50	
SIC_PIBIC_2000	SemestreIngressoo	1	
SIC_PIBIC_2000	IraAluno	3	
SIC_PIBIC_2000	MatriculaAluno	7	
SIC_PIBIC_2000	NomeAluno	50	
SIC_PIBIC_2000	Sexo	1	
SIC_PIBIC_2000	VinculoEmpregaticio	1	
SIC_PIBIC_2000	CodigoCurso	1	
SIC_PIBIC_2000	AnoIngresso	4	
SIC_PIBIC_2000	ReprovaçõesAnteriores		

SIC_PIBIC_2000	ReprovaçõesOK	1	
SIC_PIBIC_2000	RgNumeroAluno	18	
SIC_PIBIC_2000	RgOrgaoEmissorAluno	5	
SIC_PIBIC_2000	RgUfAluno	3	
SIC_PIBIC_2000	Endereco	70	
SIC_PIBIC_2000	Bairro	30	
SIC_PIBIC_2000	Cidade	50	
SIC_PIBIC_2000	Uf	3	
SIC_PIBIC_2000	Cep	12	
SIC_PIBIC_2000	TelefoneAluno	20	
SIC_PIBIC_2000	TelefoneCelularAluno	20	
SIC_PIBIC_2000	RamalAluno	10	
SIC_PIBIC_2000	AgenciaAluno	10	
SIC_PIBIC_2000	ContaAluno	15	
SIC_PIBIC_2000	InadimplenciaPendente	1	
SIC_PIBIC_2001	CpfOrientador	15	
SIC_PIBIC_2001	NomeOrientador	50	
SIC_PIBIC_2001	TitulacaoOrientador	5	
SIC_PIBIC_2001	RgNumeroOrientador	18	
SIC_PIBIC_2001	RgOrgaoEmissorOrientador	10	
SIC_PIBIC_2001	RgUfOrientador	2	
SIC_PIBIC_2001	CodUnidadeOrientador	2	
SIC_PIBIC_2001	FuncaoOrientador	5	
SIC_PIBIC_2001	Endereco	70	
SIC_PIBIC_2001	Bairro	30	
SIC_PIBIC_2001	Cidade	50	
SIC_PIBIC_2001	Uf	3	
SIC_PIBIC_2001	Cep	12	
SIC_PIBIC_2001	Telefone	20	
SIC_PIBIC_2001	TelefoneCelular	20	
SIC_PIBIC_2001	Ramal	10	
SIC_PIBIC_2001	Email	50	
SIC_PIBIC_2001	SubAreaCnpqOrientador	20	
SIC_PIBIC_2001	Ativo	1	
SIC_PIBIC_2001	DataInativo		
SIC_PIBIC_2001	OrientadorInadimplente	1	
SIC_PIBIC_2001	BolsaRemunerada	1	
SIC_PIBIC_2001	CpfAluno	15	
SIC_PIBIC_2001	DataNascimento		
SIC_PIBIC_2001	EmailAluno	50	
SIC_PIBIC_2001	SemestreIngressoo	1	

SIC_PIBIC_2001	IraAluno	3	2
SIC_PIBIC_2001	MatriculaAluno	7	
SIC_PIBIC_2001	NomeAluno	50	
SIC_PIBIC_2001	Sexo	1	
SIC_PIBIC_2001	VinculoEmpregaticio	1	
SIC_PIBIC_2001	CodigoCurso	1	
SIC_PIBIC_2001	AnoIngresso	4	
SIC_PIBIC_2001	ReproçoesAnteriores		
SIC_PIBIC_2001	ReproçoesOK	1	
SIC_PIBIC_2001	RgNumeroAluno	18	
SIC_PIBIC_2001	RgOrgaoEmissorAluno	5	
SIC_PIBIC_2001	RgUfAluno	3	
SIC_PIBIC_2001	Endereco	70	
SIC_PIBIC_2001	Bairro	30	
SIC_PIBIC_2001	Cidade	50	
SIC_PIBIC_2001	Uf	3	
SIC_PIBIC_2001	Cep	12	
SIC_PIBIC_2001	TelefoneAluno	20	
SIC_PIBIC_2001	TelefoneCelularAluno	20	
SIC_PIBIC_2001	RamalAluno	10	
SIC_PIBIC_2001	AgenciaAluno	10	
SIC_PIBIC_2001	ContaAluno	15	
SIC_PIBIC_2001	InadimplenciaPendente	1	
SIC_PIBIC_2001	NumeroPlanoTrabalho	6	
SIC_PIBIC_2001	MatriculaAluno	7	
SIC_PIBIC_2001	CpfOrientador	15	
SIC_PIBIC_2001	NumeroProjetoRelacionado	6	
SIC_PIBIC_2001	AreaPlanoTrabalho	10	
SIC_PIBIC_2001	MotivoExclusao	150	
SIC_PIBIC_2001	RecursoApresentado	1	
SIC_PIBIC_2001	PedidoReconsideracao		
SIC_PIBIC_2001	CondicaoPlanoTrabalho	20	
SIC_PIBIC_2001	TipoBolsa	20	
SIC_PIBIC_2001	FontePagadora	10	
SIC_PIBIC_2001	Palavra_chave1	30	
SIC_PIBIC_2001	Palavra_chave2	30	
SIC_PIBIC_2001	Palavra_chave3	30	
SIC_PIBIC_2001	RelatorioParcialEntregue	1	
SIC_PIBIC_2001	ApresentacaoOralOk	1	
SIC_PIBIC_2001	ResumoEntregue	1	
SIC_PIBIC_2001	RelatorioFinalEntregue	1	

SIC_PIBIC_2001	ApresentacaoCongresso	1	
SIC_PIBIC_2001	NomeAssessorExtra	50	
SIC_PIBIC_2001	AvaliacaoInserida1	1	
SIC_PIBIC_2001	AvaliacaoInserida2	1	
SIC_PIBIC_2001	Inadimplente	1	
SIC_PIBIC_2002	CpfOrientador	15	
SIC_PIBIC_2002	NomeOrientador	50	
SIC_PIBIC_2002	TitulacaoOrientador	5	
SIC_PIBIC_2002	RgNumeroOrientador	18	
SIC_PIBIC_2002	RgOrgaoEmissorOrientador	10	
SIC_PIBIC_2002	RgUfOrientador	2	
SIC_PIBIC_2002	CodUnidadeOrientador	2	
SIC_PIBIC_2002	FuncaoOrientador	5	
SIC_PIBIC_2002	Endereco	70	
SIC_PIBIC_2002	Bairro	30	
SIC_PIBIC_2002	Cidade	50	
SIC_PIBIC_2002	Uf	3	
SIC_PIBIC_2002	Cep	12	
SIC_PIBIC_2002	Telefone	20	
SIC_PIBIC_2002	TelefoneCelular	20	
SIC_PIBIC_2002	Ramal	10	
SIC_PIBIC_2002	Email	50	
SIC_PIBIC_2002	SubAreaCnpqOrientador	20	
SIC_PIBIC_2002	Ativo	1	
SIC_PIBIC_2002	OrientadorInadimplente	1	
SIC_PIBIC_2002	BolsaRemunerada	1	
SIC_PIBIC_2002	CpfAluno	15	
SIC_PIBIC_2002	EmailAluno	50	
SIC_PIBIC_2002	SemestreIngressoo	1	
SIC_PIBIC_2002	IraAluno	3	4
SIC_PIBIC_2002	MatriculaAluno	7	
SIC_PIBIC_2002	NomeAluno	50	
SIC_PIBIC_2002	Sexo	1	
SIC_PIBIC_2002	VinculoEmpregaticio	1	
SIC_PIBIC_2002	CodigoCurso	1	
SIC_PIBIC_2002	AnoIngresso	4	
SIC_PIBIC_2002	ReprovaçõesAnteriores		
SIC_PIBIC_2002	ReprovaçõesOK	1	
SIC_PIBIC_2002	RgNumeroAluno	15	
SIC_PIBIC_2002	RgOrgaoEmissorAluno	10	
SIC_PIBIC_2002	RgUfAluno	2	

SIC_PIBIC_2002	Endereco	200	
SIC_PIBIC_2002	Bairro	50	
SIC_PIBIC_2002	Cidade	50	
SIC_PIBIC_2002	Uf	2	
SIC_PIBIC_2002	Cep	12	
SIC_PIBIC_2002	TelefoneAluno	20	
SIC_PIBIC_2002	TelefoneCelularAluno	20	
SIC_PIBIC_2002	RamalAluno	10	
SIC_PIBIC_2002	AgenciaAluno	10	
SIC_PIBIC_2002	ContaAluno	15	
SIC_PIBIC_2002	InadimplenciaPendente	1	
SIC_PIBIC_2002	NumeroPlanoTrabalho	6	
SIC_PIBIC_2002	MatriculaAluno	7	
SIC_PIBIC_2002	CpfOrientador	15	
SIC_PIBIC_2002	NumeroProjetoRelacionado	6	
SIC_PIBIC_2002	TituloPlanoTrabalho		
SIC_PIBIC_2002	AreaPlanoTrabalho	10	
SIC_PIBIC_2002	MotivoExclusao	150	
SIC_PIBIC_2002	RecursoApresentado	1	
SIC_PIBIC_2002	CondicaoPlanoTrabalho	20	
SIC_PIBIC_2002	TipoBolsa	20	
SIC_PIBIC_2002	FontePagadora	10	
SIC_PIBIC_2002	Palavra_chave1	30	
SIC_PIBIC_2002	palavra_chave2	30	
SIC_PIBIC_2002	palavra_chave3	30	
SIC_PIBIC_2002	RelatorioParcialEntregue	1	
SIC_PIBIC_2002	ApresentacaoOralOk	1	
SIC_PIBIC_2002	ResumoEntregue	1	
SIC_PIBIC_2002	RelatorioFinalEntregue	1	
SIC_PIBIC_2002	ApresentacaoCongresso	1	
SIC_PIBIC_2002	NomeAssessorExtra	50	
SIC_PIBIC_2002	AvaliacaoInserida1	1	
SIC_PIBIC_2002	AvaliacaoInserida2	1	
SIC_PIBIC_2002	Inadimplente	1	
SIC_PIBIC_2003	NumeroPlanoTrabalho	6	
SIC_PIBIC_2003	MatriculaAluno	9	
SIC_PIBIC_2003	CpfOrientador	15	
SIC_PIBIC_2003	NumeroProjetoRelacionado	6	
SIC_PIBIC_2003	CondicaoPlanoTrabalho	20	
SIC_PIBIC_2003	TipoBolsa	20	
SIC_PIBIC_2003	FontePagadora	50	

SIC_PIBIC_2003	CpfOrientador	15	
SIC_PIBIC_2003	NomeOrientador	50	
SIC_PIBIC_2003	TitulacaoOrientador	5	
SIC_PIBIC_2003	RgNumeroOrientador	18	
SIC_PIBIC_2003	RgOrgaoEmissorOrientador	10	
SIC_PIBIC_2003	RgUfOrientador	2	
SIC_PIBIC_2003	CodUnidadeOrientador	2	
SIC_PIBIC_2003	FuncaoOrientador	5	
SIC_PIBIC_2003	Endereco	70	
SIC_PIBIC_2003	Bairro	30	
SIC_PIBIC_2003	Cidade	50	
SIC_PIBIC_2003	Uf	3	
SIC_PIBIC_2003	Cep	12	
SIC_PIBIC_2003	Telefone	20	
SIC_PIBIC_2003	TelefoneCelular	20	
SIC_PIBIC_2003	Ramal	10	
SIC_PIBIC_2003	Email	50	
SIC_PIBIC_2003	SubAreaCnpqOrientador	20	
SIC_PIBIC_2003	Ativo	1	
SIC_PIBIC_2003	OrientadorInadimplente	1	
SIC_PIBIC_2003	BolsaRemunerada	1	
SIC_PIBIC_2003	CpfAluno	15	
SIC_PIBIC_2003	DataNascimento		
SIC_PIBIC_2003	EmailAluno	50	
SIC_PIBIC_2003	SemestreIngressoo	1	
SIC_PIBIC_2003	IraAluno	3	2
SIC_PIBIC_2003	MatriculaAluno	7	
SIC_PIBIC_2003	NomeAluno	50	
SIC_PIBIC_2003	Sexo	1	
SIC_PIBIC_2003	VinculoEmpregaticio	1	
SIC_PIBIC_2003	CodigoCurso	1	
SIC_PIBIC_2003	AnoIngresso	4	
SIC_PIBIC_2003	ReprovaçõesAnteriores		
SIC_PIBIC_2003	ReprovaçõesOK	1	
SIC_PIBIC_2003	RgNumeroAluno	18	
SIC_PIBIC_2003	RgOrgaoEmissorAluno	5	
SIC_PIBIC_2003	RgUfAluno	3	
SIC_PIBIC_2003	Endereco	70	
SIC_PIBIC_2003	Bairro	30	
SIC_PIBIC_2003	Cidade	50	
SIC_PIBIC_2003	Uf	3	

SIC_PIBIC_2003	Cep	12	
SIC_PIBIC_2003	TelefoneAluno	20	
SIC_PIBIC_2003	TelefoneCelularAluno	20	
SIC_PIBIC_2003	RamalAluno	10	
SIC_PIBIC_2003	AgenciaAluno	10	
SIC_PIBIC_2003	ContaAluno	15	
SIC_PIBIC_2003	InadimplenciaPendente	1	
SIC_PIBIC_2005-2012	ProId	4	
SIC_PIBIC_2005-2012	PltId	4	
SIC_PIBIC_2005-2012	Matricula	10	
SIC_PIBIC_2005-2012	Situacao	7	
SIC_PIBIC_2005-2012	EdiAnoEdital	4	
SIC_PIBIC_2005-2012	orice celular	8	
SIC_PIBIC_2005-2012	orice p	8	
SIC_PIBIC_2005-2012	oridadeendereço		
SIC_PIBIC_2005-2012	oricpf	11	
SIC_PIBIC_2005-2012	escolaridade	6	
SIC_PIBIC_2005-2012	oriestadocivil	10	
SIC_PIBIC_2005-2012	orifoneres	8	
SIC_PIBIC_2005-2012	orimatricula	7	
SIC_PIBIC_2005-2012	opcni vel bolsaprodcpq	2	
SIC_PIBIC_2005-2012	oriorigem	7	
SIC_PIBIC_2005-2012	Quantidade	1	
SIC_PIBIC_2005-2012	regime	2	
SIC_PIBIC_2005-2012	orisexo	1	
SIC_PIBIC_2005-2012	siglalotacao	3	
SIC_PIBIC_2005-2012	TipoBolsa	10	
SIC_PIBIC_2005-2012	Campus	13	
SIC_PIBIC_2005-2012	EdiAnoEdital	4	
SIC_PIBIC_2005-2012	aluperingunb	5	
SIC_PIBIC_2005-2012	AluIRA	5	4
SIC_PIBIC_2005-2012	AluMatricula	10	
SIC_PIBIC_2005-2012	PltPrpId	4	
SIC_PIBIC_2005-2012	alusexo	1	
SIC_PIBIC_2005-2012	Situação1	8	
SIC_PIBIC_2005-2012	Situação2	8	
SIC_PIBIC_2005-2012	Situação3	8	
SIC_PIBIC_2005-2012	Situação4	8	
SIC_PIBIC_2005-2012	TipoVínculo	7	
SIC_PIBIC_2005-2012	Semestre	5	
SIC_PIBIC_2005-2012	PltTitulo	4	

Tabela C.5: Transformação

Database	Transformação
SIC_PIBIC_2001	remunerada/voluntária
SIC_PIBIC_2001	masculino/feminino
SIC_PIBIC_2001	sim/não
SIC_PIBIC_2001	Desnormalizar com a tabela de curso
SIC_PIBIC_2002	remunerada/voluntária
SIC_PIBIC_2002	masculino/feminino
SIC_PIBIC_2002	sim/não
SIC_PIBIC_2002	Desnormalizar com a tabela de curso
SIC_PIBIC_2003	remunerada/voluntária
SIC_PIBIC_2003	masculino/feminino
SIC_PIBIC_2003	sim/não
SIC_PIBIC_2003	Desnormalizar com a tabela de curso
SIC_PIBIC_2005-2012	inscrito/inativo
SIC_PIBIC_2005-2012	masculino/feminino