



UNIVERSIDADE ESTADUAL DO NORTE DO PARANÁ
CAMPUS LUIZ MENEGHEL

MARCELO GRACIANO DUARTE

**IMPLANTAÇÃO DE UMA FERRAMENTA DE BUSINESS
INTELLIGENCE PARA DADOS DA SECRETARIA
ACADÊMICA DA UNIVERSIDADE ESTADUAL DO
NORTE DO PARANÁ - CAMPUS LUIZ MENEGHEL**

Bandeirantes

2012

MARCELO GRACIANO DUARTE

**IMPLANTAÇÃO DE UMA FERRAMENTA DE BUSINESS
INTELLIGENCE PARA DADOS DA SECRETARIA
ACADÊMICA DA UNIVERSIDADE ESTADUAL DO
NORTE DO PARANÁ - CAMPUS LUIZ MENEGHEL**

TCC apresentado à Universidade Estadual do Norte do Paraná – *campus* Luiz Meneghel – como requisito parcial para aprovação na disciplina Projeto Final II do Curso de Sistemas de Informação.

Orientador: Prof. Me. André Luis Andrade Menolli

Bandeirantes

2012

MARCELO GRACIANO DUARTE

**IMPLANTAÇÃO DE UMA FERRAMENTA DE BUSINESS
INTELLIGENCE PARA DADOS DA SECRETARIA
ACADÊMICA DA UNIVERSIDADE ESTADUAL DO
NORTE DO PARANÁ - CAMPUS LUIZ MENEGHEL**

TCC apresentado à Universidade Estadual do Norte do Paraná – *campus* Luiz Meneghel – como requisito parcial para aprovação na disciplina Projeto Final II do Curso de Sistemas de Informação.

COMISSÃO EXAMINADORA

Prof. Me. André Luis Andrade Menolli
UENP – *Campus* Luiz Meneghel

Prof. Me. Carlos Eduardo Ribeiro
UENP – *Campus* Luiz Meneghel

Prof. Me. Glauco Carlos Silva
UENP – *Campus* Luiz Meneghel

Bandeirantes, ___ de _____ de 2012

Dedico este trabalho à minha avó Laura, que faleceu durante o período de desenvolvimento deste trabalho, pois foi uma pessoa que sempre me incentivou ao longo da vida e aos meus pais que me deram muito apoio nos momentos mais difíceis.

AGRADECIMENTOS

Agradeço primeiramente a Deus, que me deu forças para superar as dificuldades pessoais e concluir este objetivo.

Aos meus pais que acreditaram em meu potencial para finalizar este trabalho.

Ao orientador Menolli, pois me ajudou muito esclarecendo diversas dúvidas durante o desenvolvimento do trabalho sendo de fundamental importância no fechamento deste trabalho.

Aos meus amigos que me deram apoio e que me ofereceram momentos de descontração.

*Estamos afogados por
informação, mas morrendo
de fome por conhecimento.*

(John Nalsbett)

RESUMO

Este trabalho apresenta o conceito de Business Intelligence e as características das ferramentas *On-Line Analytic Processing* (OLAP) encontradas no mercado, sendo uma delas escolhida a ser implantado na Secretaria Acadêmica da UENP – Campus Luiz Meneghel. Para realizar a implantação foi seguido um conjunto de etapas para a construção de um *Data Warehouse* possibilitando gerar análises e relatórios por meio da ferramenta OLAP para ajudar os gestores na tomada de decisão e obter uma visão estratégica para melhoria dos cursos e a qualidade de ensino.

Palavras-chave: Business Intelligence, Data Warehouse, Modelagem Multidimensional, Data Mart, *On-Line Analytic Processing*

ABSTRACT

This work presents the concept of Business Intelligence and the characteristics of On-Line Analytic Processing (OLAP) tools, one of which is chosen to be implanted in the Academic Secretariat of the UENP - Luiz Meneghel college. To perform the deployment was followed by a set of steps for building a Data Warehouse allowing generate analysis and reports through the OLAP tool to help managers in decision making and get a strategic vision for improve the quality of courses and teaching.

Keywords: Business Intelligence, Data Warehouse, Multidimensional Modeling, Data Mart, On-Line Analytic Processing.

LISTA DE FIGURAS

Figura 1 Integração de Dados.....	20
Figura 2 Modelo Estrela.....	22
Figura 3 Modelo Snowflake.....	23
Figura 4 DB2 Alphablox.....	28
Figura 5 Pentaho Analysis.....	29
Figura 6 Arquitetura do Servidor OLAP Mondrian.....	30
Figura 7 OpenI Jasper.....	31
Figura 8 Oracle OLAP.....	32
Figura 9 Brain.....	33
Figura 10 Processo de Desenvolvimento da Pesquisa.....	35
Figura 11 Arquitetura de um Data Warehouse.....	38
Figura 12 Fato Matriculado.....	39
Figura 13 Fato Cancelado.....	40
Figura 14 Fato Trancado.....	41
Figura 15 Fato Formado.....	42
Figura 16 Fato Desempenho Acadêmico.....	43
Figura 17 Data Mart Integrado.....	44
Figura 18 Processo de ETL Dimensão Acadêmico.....	46
Figura 19 Select Acadêmico.....	47
Figura 20 Select Curso.....	48
Figura 21 Select Disciplina.....	48
Figura 22 Select Professor.....	49

Figura 23 ETL Dimensão Tempo.....	49
Figura 24 Script para criação de Datas.....	50
Figura 25 Select Turma	51
Figura 26 Select Curso Completo.....	51
Figura 27 ETL Fato Matriculado.....	52
Figura 28 Select Fato Matriculado.....	52
Figura 29 Script para criação de medidas Fato Matriculado.....	53
Figura 30 Select Fato Trancado.....	53
Figura 31 Script para criação de medidas Fato Trancado.....	54
Figura 32 Select Fato Cancelado.....	54
Figura 33 Script para criação de medidas Fato Cancelado.....	55
Figura 34 Select Fato Formado.....	55
Figura 35 Script para criação de medidas Fato Formado.....	56
Figura 36 Select Fato Desempenho Acadêmico.....	56
Figura 37 Script para criação de medidas Fato Desempenho Acadêmico.....	57
Figura 38 ETL Desempenho Acadêmico.....	58
Figura 39 Conexão com a base dimensional	60
Figura 40 Schema XML	61

LISTA DE ABREVIATURAS E SIGLAS

DW	Data Warehouse
ETL	Extração, Transformação e Carga (Extract Transform Load)
OLAP	On Line Analytical Processing
SGBD	Sistema Gerenciador de Banco de Dados
SQL	Structured Query Language
BI	Business Intelligence
XML	Extensible Markup Language
UENP	Universidade Estadual do Norte do Paraná
DM	Data Mart
GPL	General Public License
MDX	Multi Dimensional Expression
API	Application Programming Interface
KPI	Key Performance Indicator

SUMÁRIO

1	INTRODUÇÃO	15
1.1	JUSTIFICATIVA	17
1.2	OBJETIVO	17
1.2.1	OBJETIVOS ESPECÍFICOS	17
1.3	ORGANIZAÇÃO DO TRABALHO	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	DATA WAREHOUSE	19
2.2	ETL	20
2.3	MODELAGEM MULTIDIMENSIONAL	21
2.4	GRANULARIDADE DE DADOS	23
2.5	ARQUITETURA DE DATA WAREHOUSE	24
2.6	METADADOS	25
3	FERRAMENTAS DE BUSINESS INTELLIGENCE	26
3.1	CONCEITO DE BI	26
3.2	MICROSOFT SQL SERVER 2008 REPORT BUILDER	27
3.3	DB2 ALPHABLOX	27
3.4	PENTAHO ANALYSIS COMMUNITY	28
3.4.1	Servidor OLAP Mondrian	29
3.5	OPENI JASPER	31
3.6	ORACLE OLAP	31
3.7	BRAIN	32
3.8	RESUMO DAS CARACTERÍSTICAS DAS FERRAMENTAS	33
4	METODOLOGIA DE DESENVOLVIMENTO	34
4.1	SELEÇÃO DA BASE	35
4.2	CRIAÇÃO DO MODELO DIMENSIONAL	36
4.3	CRIAÇÃO DO MODELO FÍSICO	36
4.4	ETL	36

4.5	IMPLANTAÇÃO NA FERRAMENTA OLAP	36
5	DESENVOLVIMENTO	37
5.1	DEFINIÇÃO DO DATA MART A SER ESPECIFICADO	37
5.2	MODELAGEM DO DATA MART DIMENSIONALMENTE	37
5.2.1	Arquitetura BUS.....	37
5.2.2	Processo de Negócio	38
5.2.3	Fato Matriculado.....	38
5.2.4	Fato Cancelado	39
5.2.5	Fato Trancado	40
5.2.6	Fato Formado.....	41
5.2.7	Fato Desempenho Acadêmico	42
5.2.8	Data Mart Integrado.....	43
5.2.9	Dimensão com vários papéis	44
5.2.10	Fatos aditivos e semi aditivos.....	45
6	ETL	46
6.1	DIMENSÃO ACADÊMICO	46
6.2	DIMENSÃO CURSO	48
6.3	DIMENSÃO DISCIPLINA.....	48
6.4	DIMENSÃO PROFESSOR.....	49
6.5	DIMENSÃO TEMPO.....	49
6.6	DIMENSÃO TURMA.....	51
6.7	DIMENSÃO CURSO COMPLETO	51
6.8	FATO MATRICULADO.....	52
6.9	FATO TRANCADO	53
6.10	FATO CANCELADO	54
6.11	FATO FORMADO	55
6.12	FATO DESEMPENHO ACADÊMICO	56
6.13	CONCLUSÃO	58
7	IMPLANTAÇÃO DO DATA MART NA FERRAMENTA OLAP	59
8	ANÁLISE	61

8.1 FORMANDOS POR CURSO.....	61
9 CONCLUSÃO	63
9.1 TRABALHOS FUTUROS	63
REFERÊNCIAS.....	65

1 INTRODUÇÃO

As empresas sempre buscam melhorar seus negócios e objetivar lucros, porém evoluir é um processo complexo que às vezes depende das experiências de quem administra o negócio. Para que as decisões dos administradores não sejam feitas apenas por meio de aspectos intuitivos ou de conhecimento empírico o conceito de *Business Intelligence* teve seu espaço ampliado, pois as sucessões de decisões erradas podem comprometer um setor ou a empresa num todo, podendo levá-la à falência.

Devido a estes problemas, o conceito de *Business Intelligence* (Inteligência de Negócio) começou a ganhar maior importância. *Business Intelligence* é um conceito que amplia a utilização da informação nas organizações e dá suporte na tomada de decisão por seus administradores. Existem muitas de definições para o termo, entre elas:

O conceito de *Business Intelligence* com o entendimento de que é Inteligência de Negócios ou Inteligência Empresarial compõe-se de um conjunto de metodologias de gestão implementadas através de ferramentas de software, cuja função é proporcionar ganhos nos processos decisórios gerenciais e da alta administração nas organizações, baseada na capacidade analítica das ferramentas que integram em um só lugar todas as informações necessárias ao processo decisório. Reforça-se que o objetivo de *Business Intelligence* é transformar dados em conhecimentos, que suporta o processo decisório com o objetivo de gerar vantagens competitivas. (ANTONELLI, 2009 apud ANGELONI; REIS, 2006, p. 3)

Uma forma de implantação da tecnologia de BI é por meio de ferramentas *On-Line Analytic Processing* (OLAP), o qual apresenta análises e realizam relatórios precisos sobre os dados armazenados em bancos de dados, planilhas eletrônicas, entre outros. Estas ferramentas permitem a visualização de um determinado assunto em forma de gráficos, possibilitam por meio do *Key Performance Indicator* (KPI), verificar se algum aspecto está fora da normalidade.

O *Business Intelligence* (BI) por ser uma ferramenta de gestão pode ser aplicado a quem deseja aprimorar suas decisões, como por exemplo, Universidades, que por meio deste tipo de ferramentas poderia se ter uma visão estratégica dos cursos, alunos e o desempenho do mesmo, podendo auxiliar assim na melhoria da qualidade de ensino.

Sendo assim, neste trabalho são apresentadas algumas ferramentas OLAP existentes no mercado, e dentre todas as pesquisadas, foi escolhida a ferramenta BRAIN, para ser aplicada na Secretaria Acadêmica do Campus Luiz Meneghel – Universidade do Norte do Paraná (UENP). A escolha da ferramenta BRAIN se deu principalmente por ser uma ferramenta gratuita com idioma em português.

No entanto, a implantação de uma ferramenta OLAP não é uma tarefa trivial, pois existe todo um processo a ser seguido antes da implantação da mesma. Uma etapa importante a ser realizada é a modelagem multidimensional dos dados. Segundo Kimball (1998) é uma técnica de modelagem lógica de um projeto de banco de dados dentro de uma estrutura padrão e intuitiva permitindo ainda realizar consultas de forma ágil. Assim, este modelo ilustra um determinado processo de negócio a ser analisado. Os modelos dos bancos transacionais não são indicados para este processo, porque dificultam o entendimento do assunto a ser analisado.

A partir do modelo dimensional criado, a próxima etapa é a construção de um *Data Warehouse* (DW). Trata-se de um repositório, que faz a integração dos dados de vários locais, sendo tratados e preparados para ajudar no suporte a tomada de decisão. Segundo Centenaro et al. (S/D) o conceito de DW consiste em organizar os dados corporativos de maneira eficaz, dando suporte de informações aos administradores das empresas para as decisões de nível estratégico, com um banco de dados paralelo aos sistemas operacionais da empresa. Sua construção consiste em extrair os dados da base de origem, passar pelo processo de integração e limpeza dos dados nomeados ETL (*Extract Transform Load*) e carregá-los em outra base de destino definido na modelagem multidimensional. O DW possibilita a extração de um grande número de informações, na qual a ferramenta OLAP tem o papel de executar as análises, mostrando o cenário que a organização se encontra.

Ao construir um DW e implantar a ferramenta OLAP, a organização ganha em produtividade e tempo, pois todo o processo das informações é ágil, tendo assim um processo decisório mais rápido e com maior segurança.

1.1 JUSTIFICATIVA

Atualmente na Universidade do Norte do Paraná (UENP) – Campus Luiz Meneghel há uma grande quantidade de dados em sua base, porém eles ficam armazenados e pouco traz de conhecimento para analisar um curso ou um aluno respectivamente, isto é, os dados ficam subutilizados. Este trabalho propõe a criação de um *Data Mart* (DM) e a implantação de uma ferramenta OLAP para a Secretaria Acadêmica do Campus Luiz Meneghel, para que sejam realizadas análises por meio dos gestores para auxiliar na tomada de decisão e também conhecer a realidade do Campus. Futuramente poderá ser aplicada a construção de um DW integrando as diversas bases dos diversos sistemas existentes na UENP, oferecendo uma visão global da Universidade. Além da visão integrada que a construção de um *Data Warehouse* pode proporcionar, podem ser aplicadas outras técnicas como, por exemplo, Mineração de Dados.

1.2 OBJETIVO

O objetivo principal deste trabalho é implantar um sistema de BI OLAP na Secretaria Acadêmica da UENP (Universidade Estadual do Norte do Paraná) – Campus Luiz Meneghel, que integre os dados do departamento, a fim de dar suporte na análise destes dados.

1.2.1 OBJETIVOS ESPECÍFICOS

- Definir o *Data Mart* a ser especificado;
- Modelar o *Data Mart* dimensionalmente;
- Carregar os dados do modelo de origem para o *Data Mart*;
- Implantar o *Data Mart* na ferramenta OLAP;
- Testar e analisar os dados na ferramenta OLAP;

1.3 ORGANIZAÇÃO DO TRABALHO

Além deste capítulo, este trabalho está dividido em 9 capítulos:

No capítulo 2 é apresentada a fundamentação teórica dos principais conceitos utilizados na literatura num ambiente de Data Warehouse.

No capítulo 3 são apresentadas algumas das ferramentas de BI OLAP disponíveis no mercado.

No capítulo 4 é exposto a metodologia do trabalho.

No capítulo 5 é apresentado o início do desenvolvimento começando pela seleção da base e modelagem.

No capítulo 6 é apresentado as dimensões e os fatos carregados na etapa de ETL.

No capítulo 7 é apresentado a implantação na ferramenta por meio da ferramenta Schema Workbenck.

No capítulo 8 é apresentado a análise por meio da ferramenta OLAP.

No capítulo 9 é apresentado o encerramento dos estudos e conclusão do trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção apresenta uma visão geral de conceitos e elementos da construção de um Data Warehouse encontrada na literatura.

2.1 DATA WAREHOUSE

Um DW tem o objetivo de integrar dados de diferentes fontes e formatos. O DW não é construído para suportar o processo funcional ou operacional da empresa, ou seja, não é o fim, mas o meio para facilitar o uso da informação (MENOLLI, 2004 apud KIMBALL et al., 1998).

Esta definição de Kimball, reflete para a utilidade da informação na qual DW armazena para ajudar os tomadores de decisão, mas vale ressaltar que ele é resultado de um longo e complexo processo para sua construção. Outra explicação mais detalhada sobre DW é que se preocupa em integrar as informações de fontes internas e externas, de vários locais, organizando, filtrando e limpando os dados, sendo preparados para análise e suporte à decisão (MACHADO, 2010).

Para que o DW seja eficiente e eficaz Inmon (2005) destaca que existem quatro principais características:

- **Orientado por assunto:** armazena as informações mais importantes e de interesse da empresa para que os dados sejam analisados.
- **Variação de tempo:** representa resultados operacionais em um determinado momento de tempo em que foram capturados, seus dados são como um *snapshot*, um conjunto de dados estáticos e não podem ser atualizados.
- **Não Volátil:** os dados inseridos no DW não devem mudar, ao contrário do banco de dados operacional que muda quando a transação é processada.
- **Integração:** consiste em dar uma visão única dos dados, pois existem vários atributos diferentes e ao fazer a extração do banco de dados operacional para o DW é feita uma padronização como pode-se observar na Figura 1.

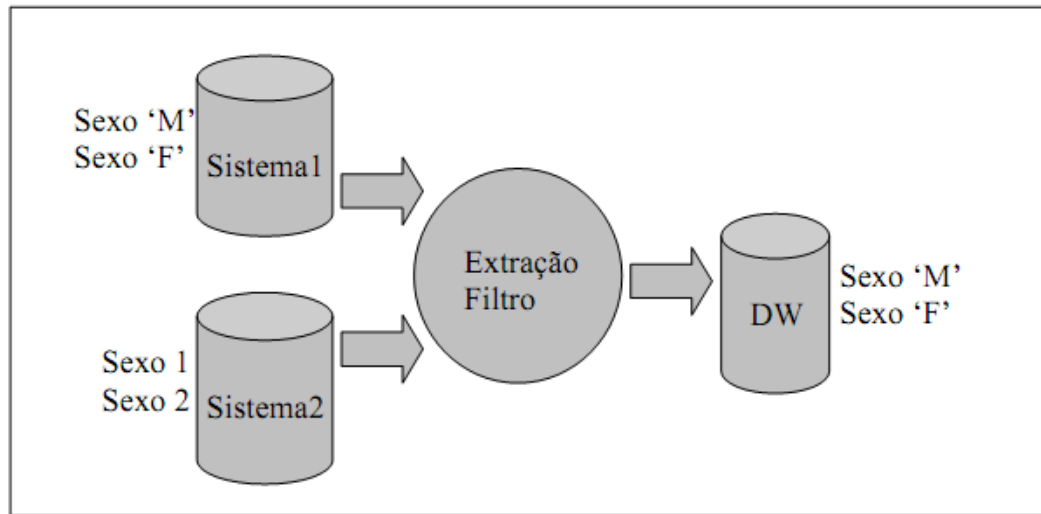


Figura 1 – Integração de Dados

Fonte: (MACHADO, 2000)

2.2 ETL

O processo de ETL (*Extract Transform Load*) é uma das etapas mais trabalhosas num projeto de DW, pois consiste em extrair os dados de uma base de origem, transformá-los e carregá-los num DW ou DM. Os dados podem ser de diversas bases o que pode dificultar o processo.

A transformação é a parte que após os dados extraídos de uma base de origem, são selecionados, limpos e tratados. Nesta etapa também é adicionado chaves substitutas que ficam invisíveis para os usuários servindo de ligação entre tabelas de dimensões e fatos.

Inmon (2005) entende que o processo de ETL pode ser automatizado ao utilizar uma ferramenta para integrar dados de várias fontes e de sistemas legados, a vantagem neste processo é que a transformação pode ser feita para grandes quantidades de dados, mas, sua desvantagem é que ao pular a etapa de transformação, ou seja, ignorá-la, a qualidade do DW fica muito reduzida.

Ao terminar de realizar a ETL, têm-se os dados filtrados no modelo

dimensional, prontos para povoar o DW e conseqüentemente ser utilizado para a realização das análises.

2.3 MODELAGEM MULTIDIMENSIONAL

Para Machado (2010), “a modelagem multidimensional é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócios”. A visualização dos dados torna-se facilitada por meio de um cubo de dados.

A modelagem dimensional para Kimball, *et al* (2008) é definido como um método de modelagem de dados para aumentar o desempenho de consultas analíticas e para facilitar seu uso, por meio de um desenho lógico para a estruturação dos dados, sendo intuitivo para os usuários de negócio.

O modelo multidimensional é formado por três elementos necessários:

- **Fatos:** é uma coleção de dados, constituído de medidas de dados e de contexto, representa uma transação e é utilizado para fazer a análise do processo de negócio de uma empresa. A tabela de fatos tem relacionamento com as tabelas de dimensão, pois está associado às dimensões por meio de chaves primárias.
- **Dimensões:** são elementos participantes de um fato, possuem descrições do negócio e determinam o assunto a ser analisado.
- **Medidas:** uma propriedade numérica que representa um fato, normalmente definido pela combinação das dimensões que fazem parte de um fato.

O modelo dimensional pode sofrer variações, porém se tem dois modelos conhecidos na literatura que é o Modelo Star ou Estrela e o Modelo Snowflake (Floco de Neve).

O Modelo Estrela é constituído de uma tabela de fato localizada no centro e um conjunto de tabelas de dimensão ao seu redor, cujas ligações são simples e de fácil entendimento. Tem como característica ser desnormalizado, aumentando a velocidade das consultas no DW. Na Figura 2 é ilustrado o modelo estrela.

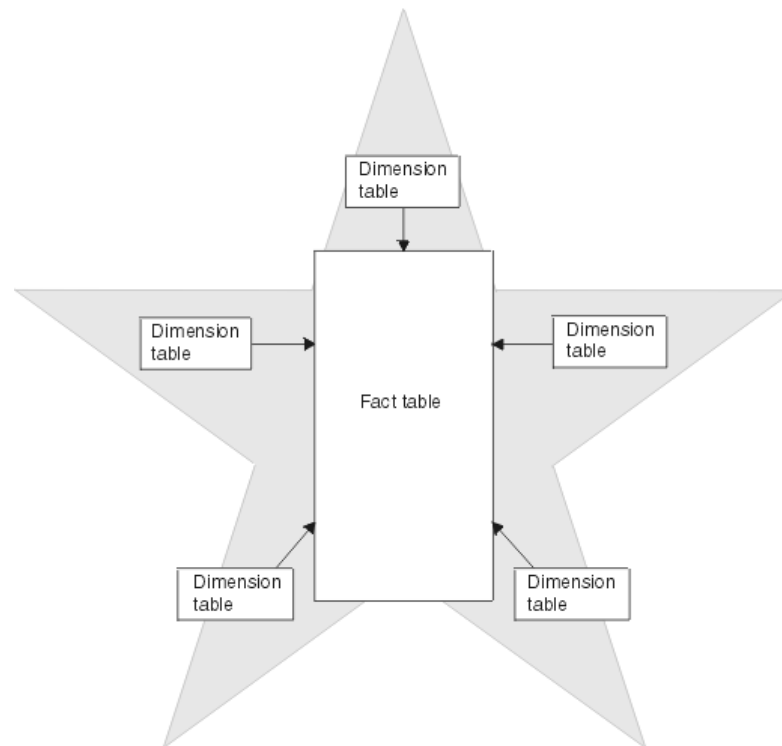


Figura 2 – Modelo Estrela

Fonte: (IBM CORPORATION, 2011)

O Modelo *Snowflake* é formado por uma tabela de fato, dimensões e subdimensões, na qual as dimensões são normalizadas, por um lado diminui o número de registros, mas por outro, as consultas são mais lentas por existir um maior número de chaves primárias e a interpretação deste modelo é mais complexa. Na Figura 3 é mostrado este modelo.

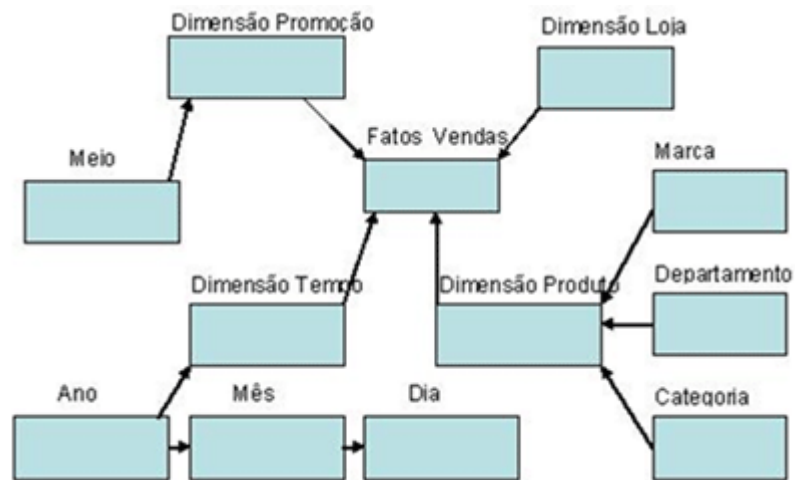


Figura 3 – Modelo Snowflake

Fonte: (MOREIRA, 2006)

2.4 GRANULARIDADE DE DADOS

Segundo Machado (2010), “a granularidade de dados refere-se ao nível de sumarização dos elementos e de detalhe disponíveis nos dados, considerando o mais importante aspecto do projeto de um Data Warehouse”.

Quanto menor é a granularidade, maior será o nível de detalhe, sendo que a granularidade maior, menor será o nível de detalhe.

Definir a granularidade é uma das etapas mais importantes no projeto de um DW, porque afeta principalmente o volume de dados a serem armazenados e também a performance das consultas.

Para Inmon (2005), a granularidade é importante para a arquitetura de DW, porque afeta a todos os ambientes relacionados.

Definir qual granularidade a ser aplicada depende do assunto a ser modelado e se já é estabelecida no começo do projeto o restante do desenvolvimento poderá se dar com mais segurança.

Ao aplicar a granularidade muito alta em determinados casos na qual se deseja armazenar dados históricos no DW, sua aplicação é válida, porém em outras situações a utilização dos dados em relatórios, poderá faltar detalhes importantes na análise.

Estabelecer a granularidade em um DM ou DW para um determinado assunto de negócios, deve ser utilizada para todo o DW onde existir referência a este assunto (MACHADO, 2010). Para realizar a análise ao encontrar diferentes granularidades poderá ocorrer alguns problemas como conflitos nas informações, dificultando o entendimento dos resultados.

2.5 ARQUITETURA DE DATA WAREHOUSE

A arquitetura de um Data Warehouse determina a forma de implantação e implementação no projeto, normalmente gerenciado por um departamento de TI, e tal escolha, pode influenciar o sucesso do projeto. A escolha da arquitetura depende da infraestrutura, recursos que será disponibilizado e também do tamanho da organização.

Existem alguns tipos de implementação de um DW, pode-se utilizar uma arquitetura global, arquitetura de DM (Data Mart) independente, arquitetura de DM integrados (MACHADO, 2010).

- **Arquitetura Global:** consiste em armazenar informações de toda a empresa num repositório comum, podendo ser fisicamente centralizada ou distribuída, além de, possuir um grande volume de acesso, por todos os departamentos. Com esta arquitetura é possível ter uma visão geral da empresa.
- **Arquitetura de DM Independente:** DM são subconjuntos ou “prateleiras” de um DW. Em relação à arquitetura trata-se de um repositório específico de dados para um determinado grupo ou departamento. Uma empresa pode ter vários DM's, porém neste caso a consulta dos dados ficará restrita a cada grupo de usuários.
- **Arquitetura de DM Integrados:** características semelhantes à arquitetura anterior, sua diferença está na integração dos vários DM's, proporcionando uma visão global da empresa. Os usuários podem acessar e utilizar os dados de um DM de outro departamento. Esta integração faz com que o nível de complexidade seja maior na sua implementação e dificultoso a sua administração.

2.6 METADADOS

Os metadados fazem parte do ambiente de DW e sua criação é de suma importância para o projeto, pois nele são contidas todas as informações sobre os dados desde a sua origem, passando pelo processo de extração até a sua utilização no DW. É um repositório de dados na qual está armazenada a documentação dos dados do banco que ajuda numa situação em que o projetista se encontra confuso no meio de tantas tabelas e relatórios.

Para Machado (2010), “os metadados não só descrevem o conteúdo de um DW, como também fornecem informações úteis para o julgamento da qualidade do conteúdo.”

Segundo Inmon (1997), os metadados incorporam o DW e mantém informações sobre o que está onde. Ele ainda define que tipos de informações os metadados mantêm:

- “A estrutura de dados, segundo a visão do programador;
- A estrutura de dados segundo a visão dos analistas de sistemas de apoio à decisão;
- A fonte de dados que alimenta o Data Warehouse;
- A transformação sofrida pelos dados no momento de sua migração para o Data Warehouse”;
- O modelo de dados;
- O relacionamento entre o modelo de dados e o Data Warehouse;
- O histórico de extrações de dados.”

3 FERRAMENTAS DE BUSINESS INTELLIGENCE

Escolher a ferramenta de BI adequada para a necessidade do projeto é uma tarefa difícil, pois o mercado está em amplo crescimento e oferece várias ferramentas, tanto pagas como gratuitas. Para determinar a ferramenta deve-se levar em consideração se ela atende as necessidades do negócio e se pode ser utilizada pelos recursos disponíveis no projeto. Neste trabalho são apresentadas algumas delas. Todas as ferramentas pesquisadas são ferramentas de BI OLAP que possuem módulos de análise gráfica, relatórios e gráficos.

3.1 Conceito de BI

Segundo Machado (2010), “BI é o conjunto de tecnologias orientadas a disponibilizar informação e conhecimento em uma empresa”. Antes as empresas tomavam decisões importantes pela intuição e percepção dos acontecimentos, muitas vezes, tinha-se um resultado inesperado por falta de confiabilidade e suporte do que realmente acontece na empresa.

Atualmente com as novas tecnologias tornou-se possível a manipulação de informações visando o conhecimento maior de sua organização, proporcionando o planejamento estratégico, tendo de seus gerentes uma visão única da empresa, com informações rápidas e precisas.

As informações podem ser encontradas em bancos de dados ou em diversas fontes de dados como planilhas eletrônicas, entre outras. O funcionamento desta tecnologia inicia-se da coleta dos dados de diversas aplicações operacionais, integrando em áreas lógicas de negócio em um repositório central, disponibilizando as informações numa ferramenta de visualização para os tomadores de decisão.

3.2 Microsoft SQL Server 2008 Report Builder

O *Microsoft SQL Server 2008 Report Builder* é um software proprietário que realiza o processamento de relatórios de alto desempenho, sendo estes oriundos de diversas fontes de dados como: SQL Server, Oracle, DB2, entre outros. Possui uma interface semelhante ao Microsoft Word, fornece visualização de gráficos e indicadores de desempenho e oferece formatação no ambiente de seus relatórios, porém por ser paga e sua aplicação ser compatível apenas para: Windows 7, Windows Server 2003 Service Pack 2, Windows Server 2008, Windows Vista Service Pack 1, Windows XP Service Pack 3, acaba se tornando uma opção pouco viável para projetos livres, de pequeno porte ou com poucos recursos.

3.3 DB2 Alphablox

O software DB2 Alphablox faz parte da suíte da *Internacional Business Machines* (IBM), é uma ferramenta OLAP para Web, ou seja, um portal na qual o usuário realiza análise dos processos de negócios em alta velocidade. Oferece uma função diferenciada ao emitir alertas visuais sobre os itens que estão fora das perspectivas normais, gráficos e tabelas intuitivas, possui interface familiar. Este sistema suporta: AIX, Linux, SCO Group (SCO) família, Solaris (Sun Microsystems), a família Windows. Este software é pago, ao adquirir a licença automaticamente se ganha o suporte por 12 meses, mas, apesar de funcionar em vários Sistemas Operacionais seu custo é elevado para este projeto. Na Figura 4 é ilustrado a interface desta ferramenta.

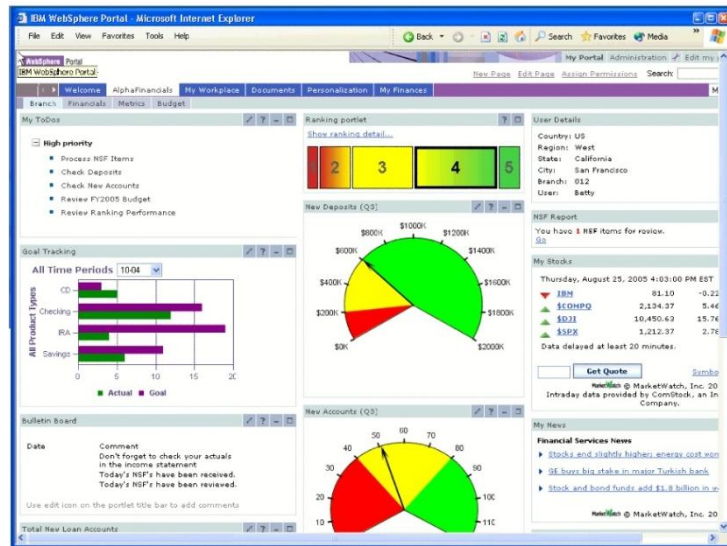


Figura 4 – DB2 Alphablox

Fonte: (IBM CORPORATION, 2011)

3.4 Pentaho Analysis Community

A Pentaho Analysis Community é uma ferramenta *open source*, que utiliza a licença GPL (*General Public Licence*), é gratuita e sua interface baseada em Web é simples o que facilita a sua utilização sem a necessidade de treinamento de usuários, as tabelas disponíveis ficam em um menu no lado direito e podem ser arrastadas ao centro da ferramenta para sua visualização e também podem ser editadas. Esta ferramenta é baseada no servidor OLAP Mondrian. Sua aplicação funciona nos seguintes sistemas operacionais: Windows, Linux e Mac. Na Figura 5 é ilustrado sua interface.

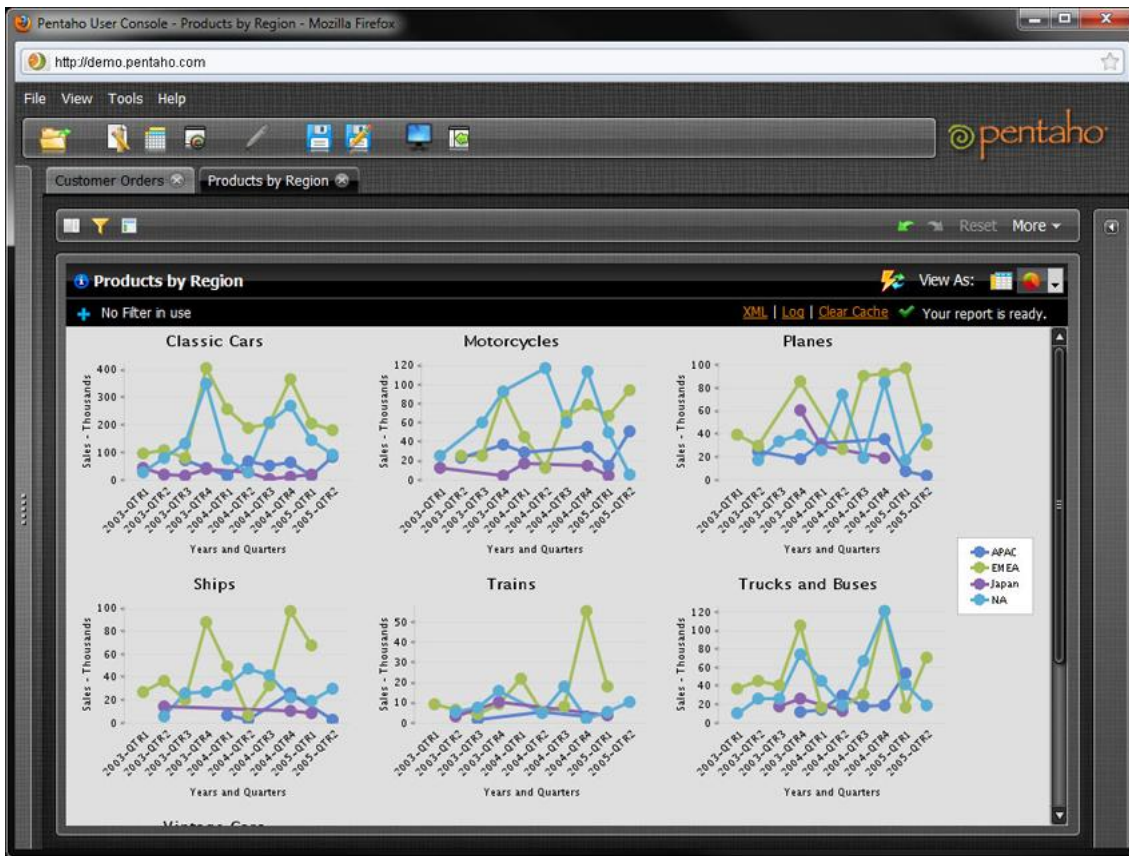


Figura 5 – Pentaho Analysis

Fonte: (PENTAHO, 2011)

3.4.1 Servidor OLAP Mondrian

O Mondrian é um servidor OLAP *open source* da Pentaho, gratuito e está apto a rodar no servidor Web Apache Tomcat. Ele é responsável por executar e retornar as consultas feitas pela ferramenta OLAP. Para visualizar as consultas o Mondrian necessita de outra ferramenta, por exemplo, o JPivot (PENTAHO, 2011).

Este servidor é responsável por executar as consultas escritas na linguagem MDX (*Multi Dimensional Expression*) convertendo para SQL, lendo os dados das bases relacionais e apresentando os dados no formato multidimensional por meio de uma API (*Application Programming Interface*) Java (PENTAHO, 2011).

O sistema Mondrian é composto por quatro camadas: camada de apresentação, camada dimensional, camada estrela e camada de armazenamento.

Assim sendo, a camada de apresentação não pertence a arquitetura do Mondrian, sendo necessário alguma ferramenta nesta camada, como por exemplo, o JPivot. A segunda camada, dimensional, é a camada que recebe, analisa, e envia consultas OLAP. A camada estrela é responsável pela ligação da consulta recebida pela camada anterior e a execução da mesma sobre a base de dados por meio de um mapeamento. E por último, a camada de armazenamento é quem executa as consultas solicitadas pela camada estrela sobre o DW (Prass *et al*, 2009, p. 56).

Na Figura 6 é ilustrada a arquitetura do Servidor Mondrian:

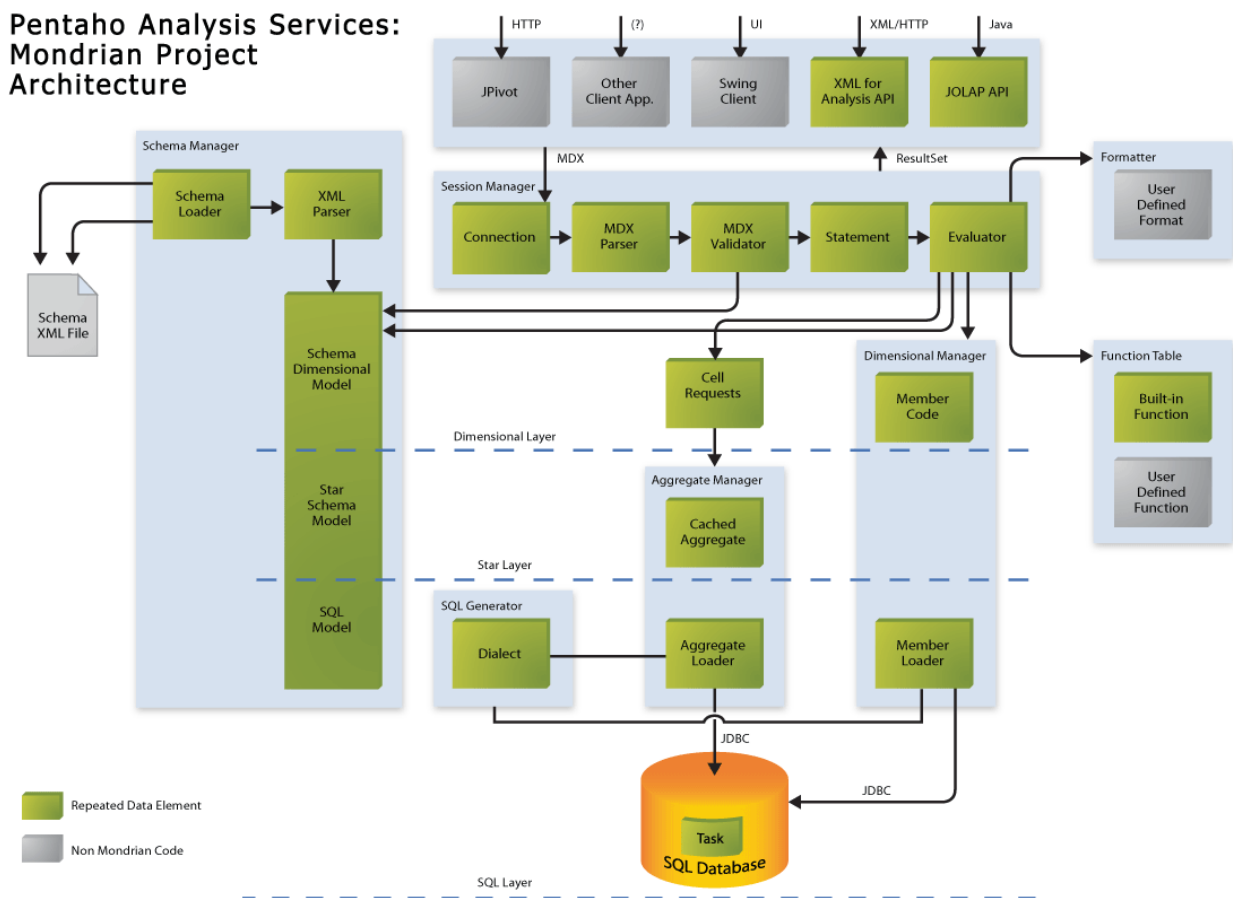


Figura 6 - Arquitetura do Servidor OLAP Mondrian

Fonte: (PENTAHO, 2011)

3.5 Open Jasper

O Jasper da OpenI é uma ferramenta *open source*, lançado em fevereiro de 2011 que possibilita aos usuários publicar relatórios novos interativos intuitivamente, seu idioma é em inglês assim como é mostrado na Figura 7. Este sistema também utiliza como base o servidor OLAP Mondrian.

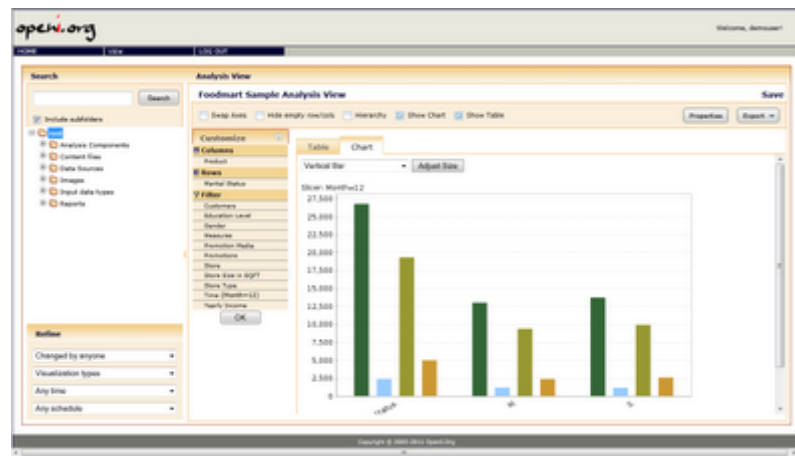


Figura 7 – OpenI Jasper

Fonte: (OPENI, 2011)

3.6 Oracle OLAP

A Oracle OLAP é um pacote com um conjunto de ferramentas para realizar todo o processo de DW desde o início ao realizar a extração dos dados da origem até o resultado final que seriam os relatórios, de característica proprietária, tem o custo de licença mais o suporte inicial à parte, este software promete realizar consultas analíticas complexas com um alto desempenho. Esta ferramenta é mais apropriada para grandes projetos, na qual o seja possível ter o retorno do investimento feito. Na Figura 8 é ilustrada sua interface.



Figura 9 – Brain

Fonte: (BRAIN, 2011)

3.8 Resumo das características das Ferramentas

Após apresentar as ferramentas de análise disponíveis, tem-se no Quadro 1 abaixo uma breve descrição de suas principais características.

Quadro 1 - Características das ferramentas OLAP

FORNECEDOR	FERRAMENTA	S.O.	SUÍTE (S/N)	TIPO	PAGA (S/N)
Microsoft	Report Builder	Windows	SIM	Desktop	SIM
IBM	DB2 Alphablox	Linux, Solaris, Windows	SIM	Web	SIM
Pentaho Community	Pentaho Analysis	Windows, Linux, Mac	NÃO	Web	NÃO
Openl	Openl Jasper	Sem restrições	NÃO	Web	NÃO
Oracle	Oracle OLAP	Sem restrições	SIM	Desktop	SIM
Brain	Brain	Sem restrições	NÃO	Web	NÃO

*S.O.: Sistema Operacional

4 METODOLOGIA DE DESENVOLVIMENTO

A metodologia deste trabalho reuniu informações no estudo da tecnologia de *data warehousing* (arquitetura, modelagem dimensional, ETL, tipos de implementação) e selecionar a ferramenta de BI OLAP mais adequada a ser implantada para fins de geração de relatórios sobre a base de dados da Divisão Acadêmica do Campus Luiz Meneghel.

Estas informações foram utilizadas a fim de criar um modelo dimensional que forneça uma visão mais ampla dos cursos e alunos, de modo a contribuir no melhor andamento do Campus.

Esta pesquisa pode ser classificada como pesquisa aplicada. Segundo Barros e Lehfeld (2000, p.78), a motivação da pesquisa aplicada é a produção do conhecimento obtido para aplicação dos seus resultados com o objetivo de “contribuir para fins práticos, visando a solução mais ou menos imediata do problema encontrado com a realidade”.

Para o desenvolvimento da pesquisa deste trabalho foram consideradas as principais etapas para a implantação da ferramenta OLAP, entre elas: seleção da base, criação do modelo dimensional, criação do modelo físico, ETL, implantação na ferramenta OLAP e as análises que são o resultado de todo este processo. Na Figura 10, no seu lado esquerdo são ilustradas as etapas para a implantação e à direita os elementos ou ferramentas envolvidas em cada etapa.

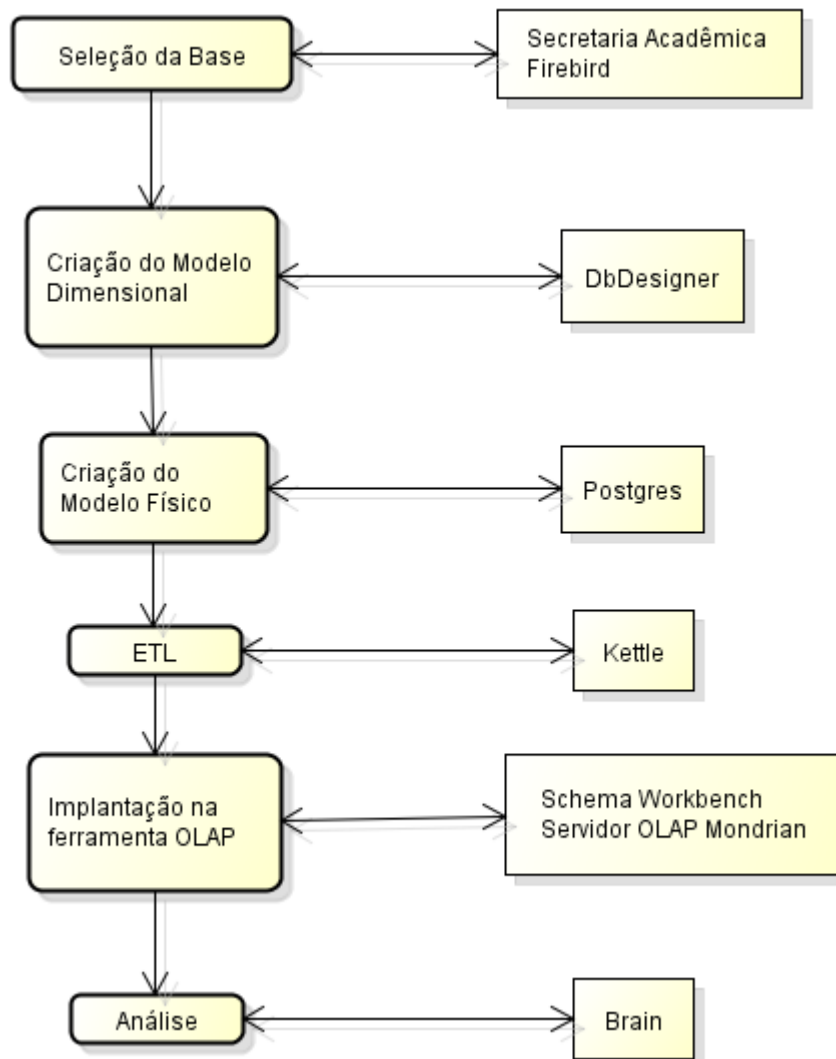


Figura 10 – Processo de Desenvolvimento da Pesquisa
Adaptado MENOLLI (2004)

4.1 Seleção da Base

A seleção da base é determinada a partir dos dados que nela estão inseridos, que possibilitem uma análise que traga valor ao Campus Luiz Meneghel e esteja diretamente relacionado com os alunos e professores. Estes que são variáveis que influenciam a qualidade dos cursos. Assim sendo, a base escolhida foi a Secretaria Acadêmica, pois reúnem todas essas condições.

4.2 Criação do Modelo Dimensional

Para a criação do modelo dimensional foi estudada a base da Secretaria Acadêmica, para a identificação de elementos a serem analisados para auxiliar no processo decisório pelos gestores do Campus Luiz Meneghel.

4.3 Criação do Modelo Físico

O modelo físico foi implementado no Postgres, pois é um SGBD *open source* e que atende as necessidades deste trabalho.

4.4 ETL

Este processo visa integrar os dados, limpá-los de forma que fiquem organizados para ser aproveitado, carregá-los no DM e, conseqüentemente, aumentar a eficiência da análise destes dados.

Para realizar a etapa de ETL foi utilizada a ferramenta *open source* Kettle da Pentaho, pois ela é gratuita e atende as necessidades deste projeto. Ela permite realizar os processos de transformações por meio dos vários componentes disponibilizados. Estes fazem a extração e o tratamento dos dados da base acadêmica que utiliza o SGBD Firebird para o SGBD Postgres, que está implementado na forma dimensional.

4.5 Implantação na ferramenta OLAP

A etapa seguinte é a implantação do *Data Mart* para que a ferramenta OLAP realize as análises de acordo com os modelos dimensionais. Primeiramente foi feito um esquema em XML (*Extensible Markup Language*) que mapeia os cubos, dimensões, hierarquias, entre outros, para que o servidor Mondrian execute as consultas OLAP e retorne os dados manipulados.

5 DESENVOLVIMENTO

5.1 Definição do Data Mart a ser especificado

A base da Secretaria Acadêmica foi escolhida, dentre os vários departamentos do Campus Luiz Meneghel, o que apresenta mais dados importantes relacionados a acadêmicos, docentes e cursos, para aplicação da ferramenta de BI OLAP. Dessa forma, o conteúdo desta base permite realizar análises em pontos críticos do Campus Luiz Meneghel, a fim de dar suporte e confiabilidade na decisão dos gestores e trazer benefícios.

A base está implementada no SGBD Firebird e é composta por 139 tabelas.

5.2 Modelagem do Data Mart Dimensionalmente

5.2.1 Arquitetura BUS

Para implantar a ferramenta OLAP na Secretaria do Campus Luiz Meneghel foi seguido o processo de arquitetura de DW que primeiro se inicia na modelagem multidimensional, em seguida é feito a extração, filtragem e limpeza dos dados da base acadêmica conhecido como a etapa de ETL realizando a carga na base dimensional que fornecerá os dados para análise na ferramenta OLAP Brain. Na Figura 11 é apresentada a arquitetura.

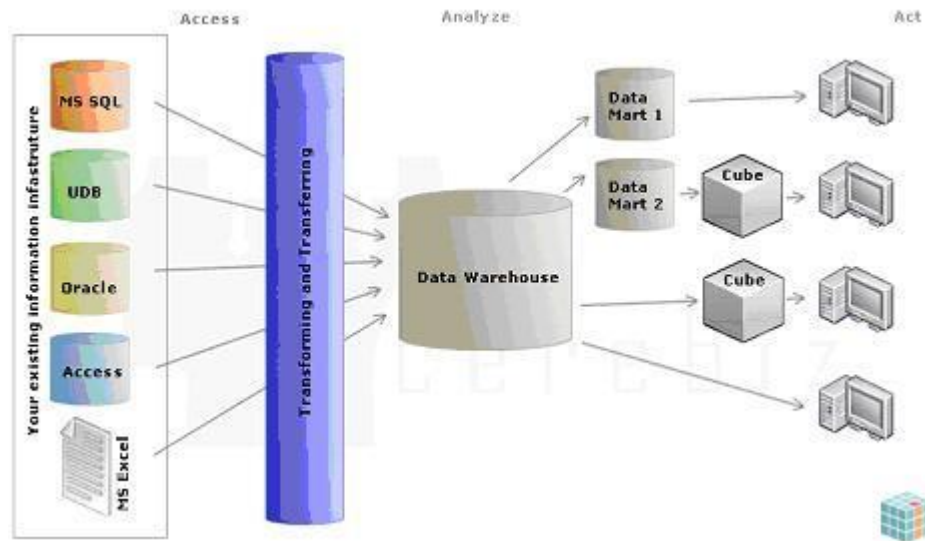


Figura 11 - Arquitetura de um Data Warehouse
(SERAIN, 2007)

5.2.2 Processo de Negócio

Para modelar dimensionalmente o DM é importante definir o processo de negócio, pois assim determina quais fatos serão analisados no Campus Luiz Meneghel. A partir do estudo e conhecimento da base de dados ficou determinado que a qualidade de ensino refletirá o andamento e a evolução dos cursos para análises dos gestores.

5.2.3 Fato Matriculado

Este fato pode medir o número de alunos matriculados em determinado curso em um período de tempo, modelado de acordo com o processo de negócio a fim de verificar a demanda por curso. O modelo é composto pelas dimensões: acadêmico, curso e tempo e a medida, total matriculado, conforme Figura 12.

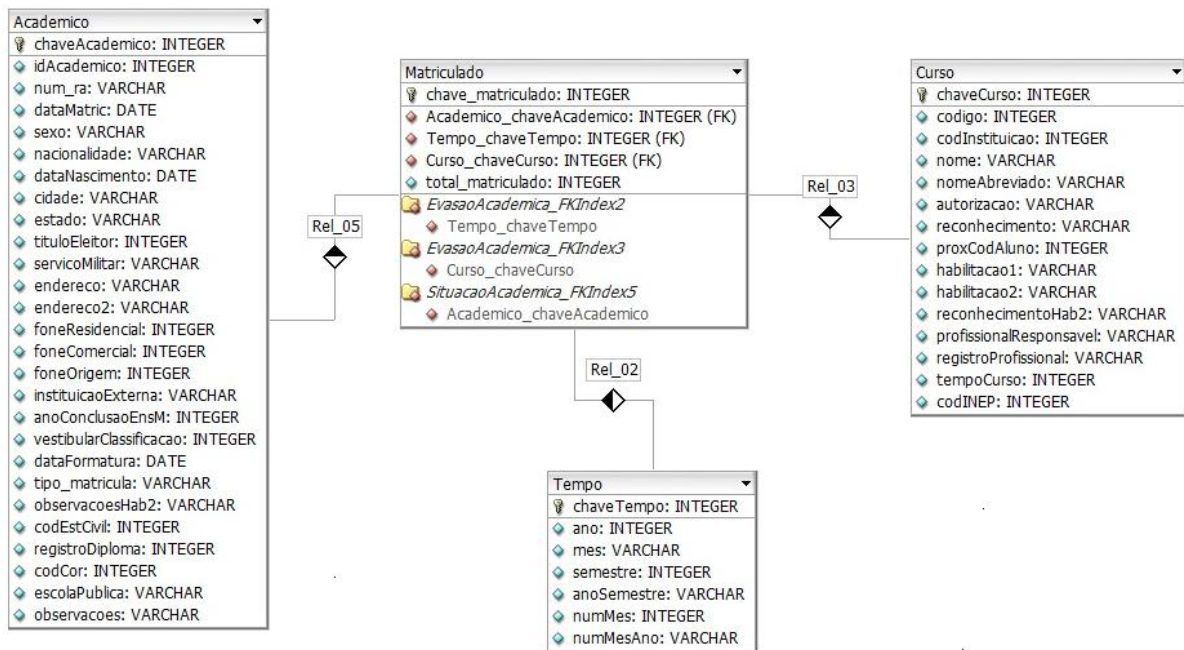


Figura 12 – Fato Matriculado

5.2.4 Fato Cancelado

Neste fato é possível verificar a quantidade de alunos que cancelaram sua matrícula e também o tempo médio cursado. Este fato contém as dimensões: acadêmico, curso, tempo e tempo final e as medidas: total cancelado e tempo cursado, como são apresentados na Figura 13.

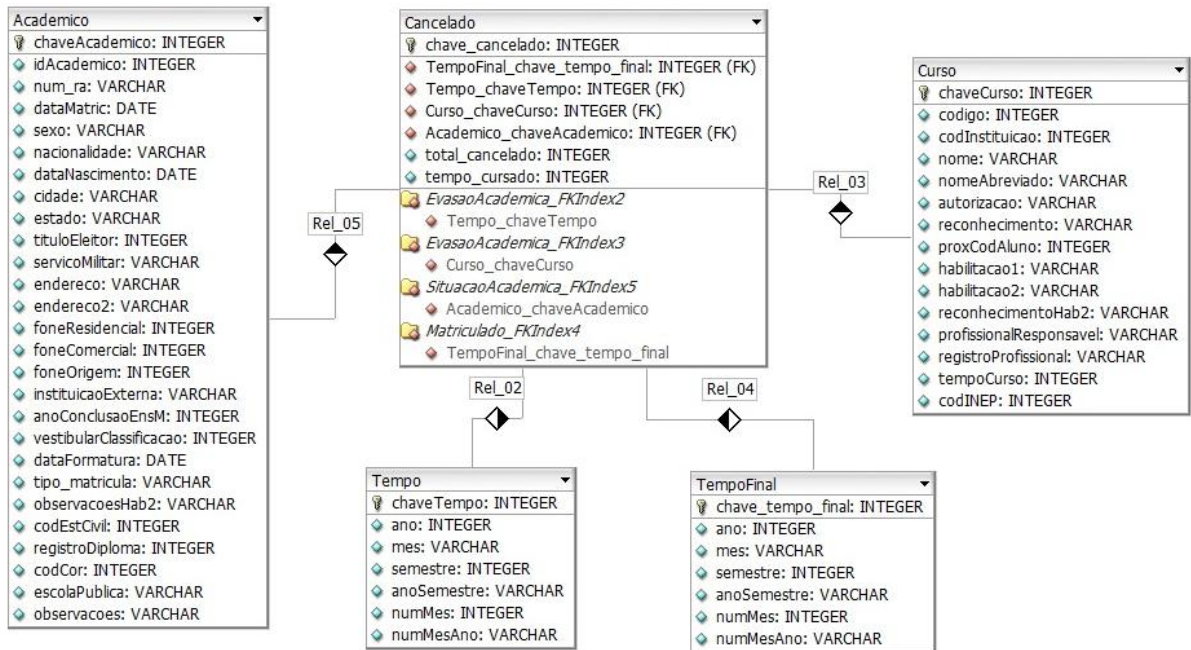


Figura 13 – Fato Cancelado

5.2.5 Fato Trancado

Neste fato é possível verificar os alunos que trancaram sua matrícula e o tempo médio cursado, composto pelas dimensões: acadêmico, curso, tempo e tempo final e medidas: total trancado e tempo cursado, conforme é mostrado na Figura 14.

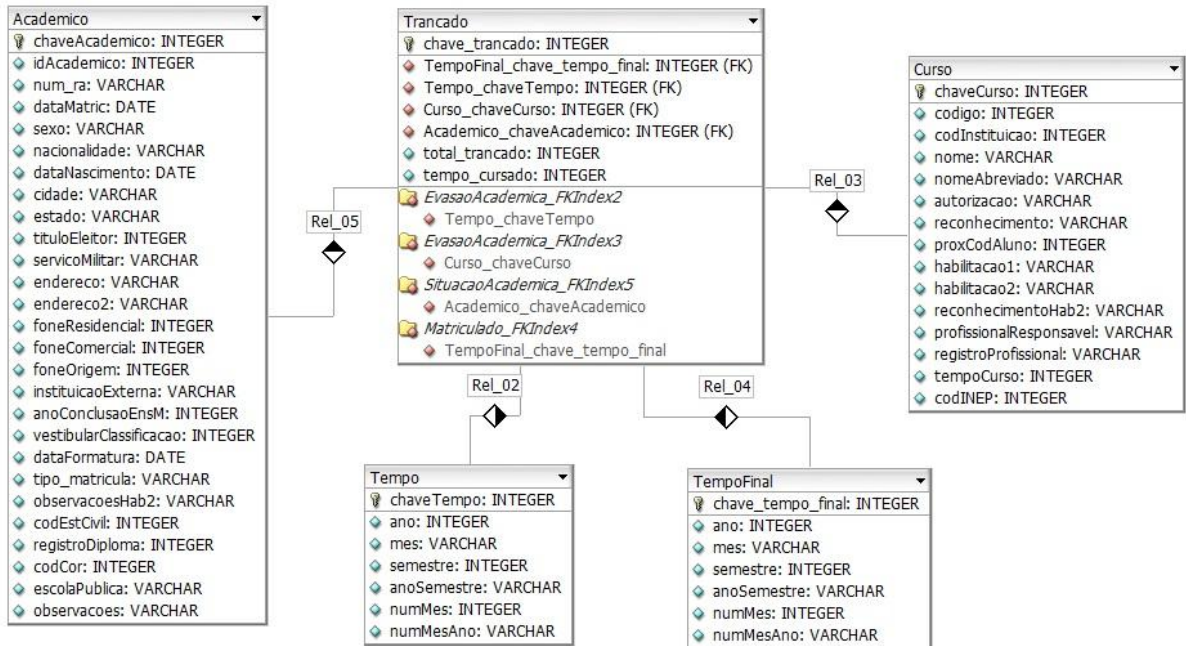


Figura 14 – Fato Trancado

5.2.6 Fato Formado

Este fato se tem a possibilidade de medir a quantidade de alunos que se formaram nos cursos do Campus Luiz Meneghel, logicamente relacionado ao processo de negócio, sendo este fato composto pelas dimensões: acadêmico, curso, tempo, tempo final e as medidas: total formado e tempo cursado, conforme o modelo dimensional mostrado na Figura 15.

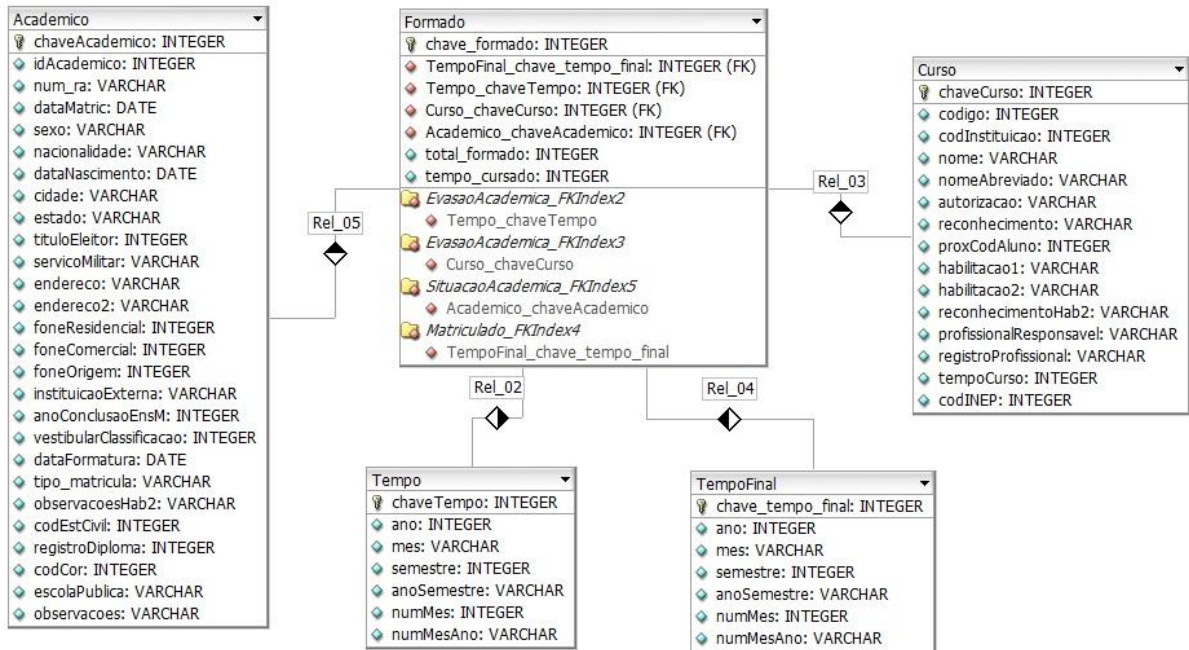


Figura 15 – Fato Formado

5.2.7 Fato Desempenho Acadêmico

Este fato tem como característica medir o desempenho dos alunos por meio das notas para isso foi feito o estudo da base acadêmica de forma a descobrir as dimensões e medidas necessárias para elaborar o modelo. Após o estudo pode-se encontrar as dimensões: curso, curso completo, professor, disciplina, turma, tempo e acadêmico, sendo as medidas: freqüência, media, media final, exame, aprovado, reprovado, aprovado no exame, reprovado no exame, reprovado sem exame, como é apresentado na Figura 16. Este modelo configura a possibilidade de medir o desempenho acadêmico de várias formas, ressaltando que a dimensão tempo neste modelo é diferente do fato anterior, pois neste o nível de detalhe é por semestre.

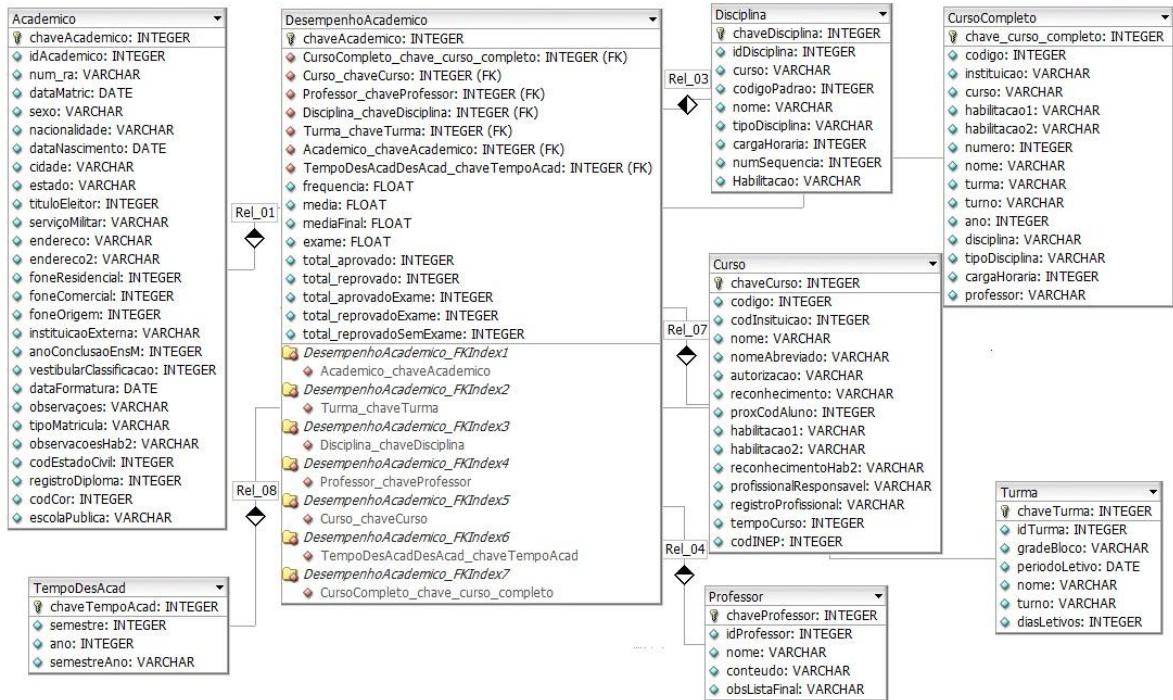


Figura 16 – Fato Desempenho Acadêmico

5.2.8 Data Mart Integrado

Este DM se trata da junção dos fatos gerados e que foi implantado na ferramenta OLAP para a Secretaria Acadêmica, dessa forma, proporciona uma melhor visão global do campus do que implementá-lo de forma separada como é mostrado na Figura 17. Os modelos compartilham algumas dimensões como, por exemplo, acadêmico, tempo e curso.

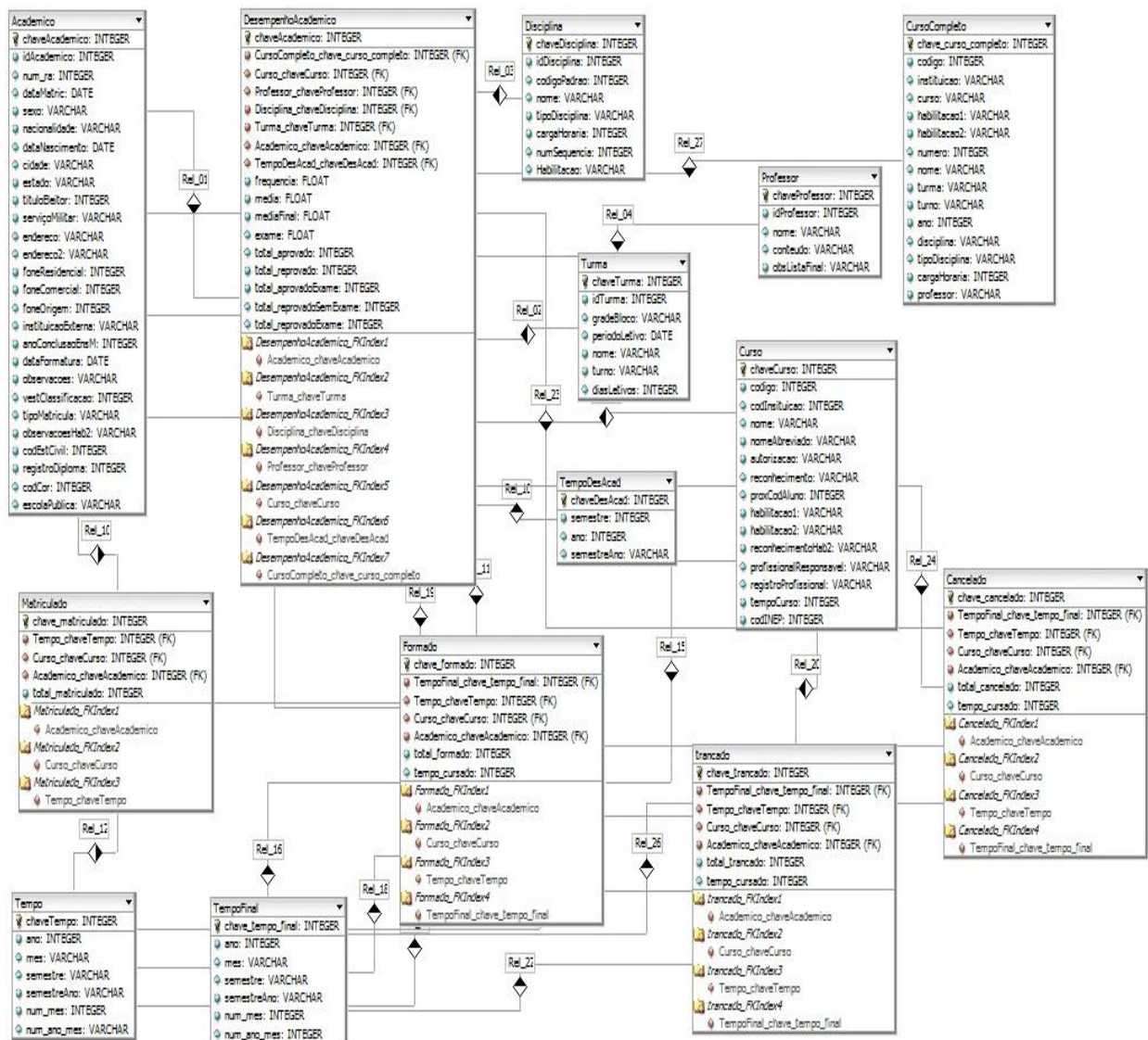


Figura 17 – Data Mart Integrado

5.2.9 Dimensão com vários papéis

Ao realizar a modelagem dimensional podem-se encontrar algumas dimensões com vários papéis, pois se trata de uma única dimensão que pode aparecer várias vezes na mesma tabela de fatos como, por exemplo, a dimensão tempo. Este problema pode ter solução com a criação de tabelas diferentes com o conteúdo semelhante.

A dimensão tempo que está relacionado ao fato evasão, por exemplo, pode ter duas ou mais datas relacionadas à mesma tabela como data de matrícula e data de trancamento. Cada um destes itens deveria estar ligado a uma dimensão tempo igual, mas que seriam tabelas diferentes para ser possível separar as colunas quando várias delas forem arrastadas para um relatório.

5.2.10 Fatos aditivos e semi aditivos

Fatos semi aditivos são medidas que não podem ser agregadas a todas as dimensões do modelo como no fato desempenho acadêmico, por exemplo, as medidas: media, media final, exame, frequência, aprovado, reprovado, aprovado no exame, reprovado no exame, reprovado sem exame não fazem sentido relacionar em pelo menos uma dimensão do modelo.

Fatos aditivos podem ser somados e relacionados às dimensões existentes no modelo como é o caso do fato evasão acadêmica que possui as seguintes medidas: matriculado, trancado, cancelado, formado, transferido, não renovou matrícula e jubilado, que é composto pelas dimensões: tempo, curso e acadêmico.

6 ETL

O processo de ETL consiste em extrair, filtrar e carregar os dados da base de origem que se encontra no SGBD Firebird para a base dimensional criada fisicamente no SGBD Postgres conforme o modelo dimensional, sendo feito apenas a carga inicial dos dados, pois para se ter periodicidade é necessário realizar a carga semestralmente quando os dados da Secretaria Acadêmica são atualizados. A ferramenta utilizada para o processo foi o Kettle, pois possui os componentes necessários para reproduzir esta etapa. O modelo dimensional possui 5 tabelas fato e 9 dimensões, sendo a dimensão tempo desempenho acadêmico criada de forma manual, por ser uma tabela com poucos registros não houve a necessidade de realizar o ETL. Para iniciar o processo foi necessário carregar primeiramente os dados das dimensões e após isso os fatos. Neste capítulo são apresentados todos os processos de ETL.

6.1 Dimensão Acadêmico

Nesta dimensão encontram-se os dados referentes a todos os alunos que estudaram ou ainda estudam no campus Luiz Meneghel, pois esta é a dimensão com mais dados a serem carregados. Informações como nome, senha e alguns documentos pessoais foram preservados filtrando apenas os dados que possibilitem realizar análises sobre o aluno. Na Figura 18 são apresentados os componentes utilizados para a extração dos dados.

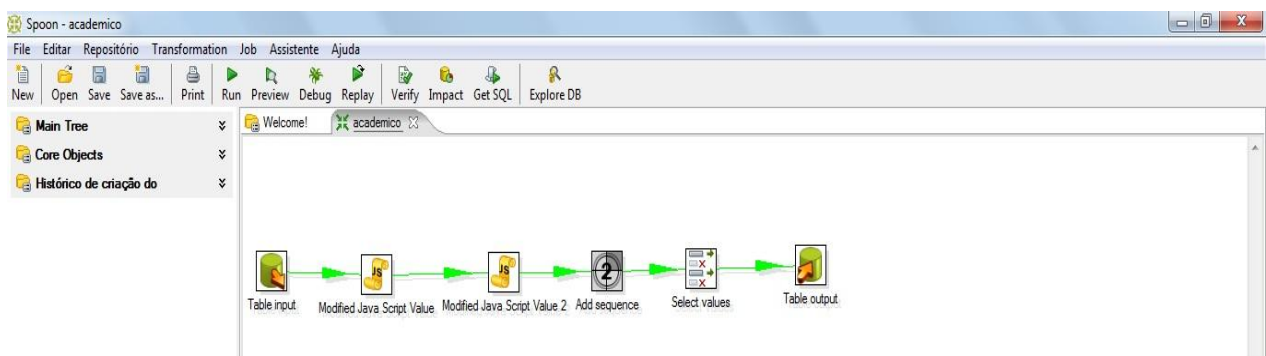


Figura 18 - Processo de ETL Dimensão Acadêmico

O componente *table input* realiza a conexão com a base de origem e seleciona os dados necessários. Sua origem é da tabela *aluno* no qual é constituída de várias chaves estrangeiras tendo a necessidade concatenar alguns campos para se adequar ao campo da dimensão que será transposta. Como é apresentado na Figura 19 o comando `SELECT`.

```
SELECT      A.OID,      A.CODIGO,      A.NUMRA,      (LPAD(extract      (month      from
A.DATAMATRICULA),2,'0')|| '/' || extract(year from a.datamatricula)) as
DATAMATRICULA, A.SEXO, N.nome AS NACIONALIDADE, A.DATANASCIMENTO, C.nome AS
CIDADE, u.nome as estado, te.numero AS TITULOELEITOR, SM.numero AS
SERVICOMILITAR, e.rua||'-'||e.numero AS ENDERECOATUAL, EN.rua||'-'||en.numero
as ENDERECOORIGEM, T.numero as FONERESIDENCIAL, TF.numero AS FONECOMERCIAL,
TEL.numero AS FONEORIGEM, IE.nome AS INSTITUICAOEXTERNA, A.ANOCONCLUMEDIO,
VA.vestclassificacao AS VESTIBULARALUNO, (LPAD(extract (month from
f.dataconclusao),2,'0')|| '/' || extract(year from f.dataconclusao)) AS
FORMATURA, A.OBSERVACOES, TM.nome AS MATRICULAALUNO, A.OBSERVACOES_HAB2,
A.CODESTADOCIVIL, RD.anoconcbacharel||'-'||anoconcllicenc as REGISTRODIPLOMA,
A.CODCOR, A.ESCOLAPUBLICA
FROM ALUNO A LEFT OUTER JOIN NACIONALIDADE N ON A.CODNACIONALIDADE = N.CODIGO
LEFT OUTER JOIN CIDADE C ON A.CODCIDADENASCIMENTO = C.CODIGO
LEFT OUTER JOIN UF U ON C.coduf=u.codigo
LEFT OUTER JOIN TITULOELEITOR TE ON A.CODTITULOELEITOR = TE.CODIGO
LEFT OUTER JOIN SERVICOMILITAR SM ON A.CODSERVICOMILITAR = SM.CODIGO
LEFT OUTER JOIN ENDERECO E ON A.CODENDERECOATUAL = E.CODIGO
LEFT OUTER JOIN ENDERECO EN ON A.CODENDERECOORIGEM = EN.CODIGO
LEFT OUTER JOIN TELEFONE T ON A.CODFONERESIDENCIAL = T.CODIGO
LEFT OUTER JOIN TELEFONE TF ON A.CODFONECOMERCIAL = TF.CODIGO
LEFT OUTER JOIN TELEFONE TEL ON A.CODFONEORIGEM = TEL.CODIGO
LEFT OUTER JOIN INSTITUICAOEXTERNA IE ON A.CODESTABMEDIO = IE.CODIGO
LEFT OUTER JOIN VESTIBULARALUNO VA ON A.CODVESTALUNO = VA.CODIGO
LEFT OUTER JOIN FORMATURA F ON A.CODFORMATURA = F.CODIGO
LEFT OUTER JOIN TIPOMATRICULAALUNO TM ON A.OIDTIPOMATRICULA = TM.OID
LEFT OUTER JOIN REGISTRODIPLOMA RD ON A.CODREGISTRODIPLOMA = RD.CODIGO;
```

Figura 19 - Select Acadêmico

Após a filtragem dos dados é colocado dois componentes javascript que tem a função de determinar o estado civil e a cor dos acadêmicos de forma textual, pois na base acadêmica está apenas representado por números, em seguida, o componente *add sequence* insere um campo na tabela acadêmico que será a chave primária desta dimensão. No passo seguinte o componente *select values*, tem a função de mapear os dados, ou seja, direcionar cada campo no seu devido local. E, por último, o componente *table output*, responsável em realizar a conexão com a base dimensional direcionando para a tabela acadêmico do Postgres. Todas as etapas são realizadas em sequência,

pois os componentes do Kettle carregam os dados do componente anterior, conforme suas ligações.

6.2 Dimensão Curso

Esta dimensão contém dados sobre os cursos de graduação do Campus Luiz Meneghel, se trata de uma tabela com poucos registros, porém de grande importância para distinção dos cursos possibilitando verificar a habilitação. O processo de ETL utiliza os componentes *table input*, *add sequence*, *select values* e *table output*, porém para selecionar os dados é realizado por meio do comando SELECT, como é mostrado na Figura 20.

```
SELECT *  
FROM CURSO C, INSTITUICAO I  
WHERE C.CODINSTITUICAO = I.CODIGO;
```

Figura 20 – Select Curso

6.3 Dimensão Disciplina

A seleção dos dados da tabela disciplina contém cadastradas todas as disciplinas dos diversos cursos e possui apenas uma chave estrangeira realizado no comando SELECT como é mostrado na Figura 21.

```
SELECT *  
FROM DISCIPLINA D, TIPODISCIPLINA T  
WHERE D.codtipodisciplina = T.codigo;
```

Figura 21 – Select Disciplina

Em seguida é realizado o processo de ETL utilizando os componentes *table input*, *add sequence*, *select values*, *table output* e inserção na base dimensional.

6.4 Dimensão Professor

Para a criação da dimensão professor os dados foram selecionados da tabela professor na base de origem com o comando SELECT como é apresentado na Figura 22.

```
SELECT * FROM PROFESSOR;
```

Figura 22 – Select Professor

Na sequência utilizou-se os componentes *table input*, *add sequence*, *select values* e *table output* para extração e carga na tabela dimensional.

6.5 Dimensão Tempo

Para construir a dimensão tempo foi utilizado um componente na qual geram de forma automática as datas necessárias para se adequar a base acadêmica. Primeiro foi necessário verificar a granularidade, neste caso, ficou definido que as datas deveriam ser geradas por mês com início em janeiro de 1977 até dezembro de 2012, pois as datas da base estão inseridas neste intervalo e assim também é possível verificar a evasão com maior detalhe. Na Figura 23 são apresentados os componentes para execução de tal processo.

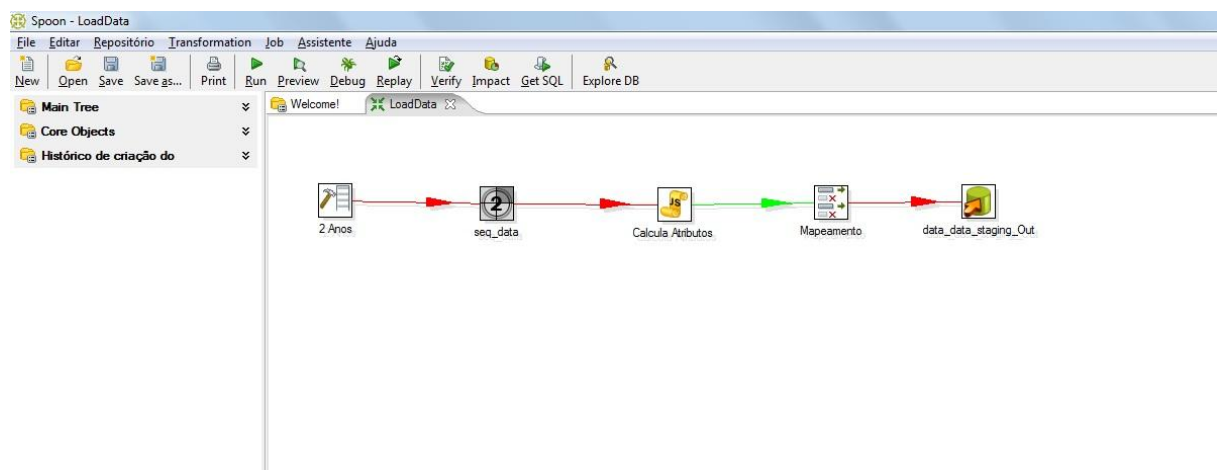


Figura 23 – ETL Dimensão Tempo

O componente `seq_data` insere uma chave primária na dimensão e o script “calcula atributos” possui o seguinte código para criação dos campos `mês`, `semestre`, `ano_ semestre`, `ano4` e `num_mes` como é mostrado na Figura 24.

```

var locale = new
java.util.Locale(language_code.getString(),country_code.getString());

var calendar = new java.util.GregorianCalendar(locale);
calendar.setTime(initial_date.getDate());

calendar.add(calendar.MONTH,DaySequence.getInteger()-1);

var data = new java.util.Date(calendar.getTimeInMillis());

var data_curta =
java.text.DateFormat.getDateInstance(java.text.DateFormat.SHORT,locale).format
(data);

var simpleDateFormat = java.text.SimpleDateFormat("D",locale);

simpleDateFormat.applyPattern("MM");
var num_mes = simpleDateFormat.format(data);

simpleDateFormat.applyPattern("MMMM");
var mes = simpleDateFormat.format(data);

simpleDateFormat.applyPattern("yyyy");
var ano4 = simpleDateFormat.format(data);

var semestre = "S";
var num_semestre;
switch(parseInt(num_mes)){
    case 1: case 2: case 3: case 4: case 5: case 6: num_semestre = "1";
break;
    case 7: case 8: case 9: case 10: case 11: case 12: num_semestre = "2";
break;
}
semestre = num_semestre;

var ano_ semestre = semestre + "-" + ano4;
var num_mes_ ano = num_mes + "/" + ano4;

```

Figura 24 – Script para criação de Datas

Após a criação dos campos, o componente *select values* mapeia para serem inseridos na tabela dimensional por meio do *table output*. Todo esse processo de ETL é repetido para a dimensão Tempo Final sendo a única diferença é com relação ao

direcionamento para a tabela correspondente na base dimensional realizado pelo componente *table output*.

6.6 Dimensão Turma

Para criar a dimensão turma foi necessário buscar os dados na base de origem turma que está composta por chaves estrangeiras, ficando o SELECT apresentado na Figura 25 do componente *table input*.

```
SELECT *
FROM TURMA T, GRADEBLOCO G, PERIODOLETIVO P, TURNOAULA TA
WHERE T.CODGRDBLOCO = G.CODIGO AND T.CODPERLETIVO = P.CODIGO AND
T.CODTURNOAULA = TA.CODIGO;
```

Figura 25 – Select Turma

O processo segue com os componentes *add sequence*, *select values* e *table output* para carregar os dados na base dimensional.

6.7 Dimensão Curso Completo

Para criar esta dimensão foi necessário selecionar alguns campos das dimensões curso, turma, disciplina e professor, como é mostrado na Figura 26.

```
SELECT td.codigo, i.sigla as instituicao, c.nome as Curso, c.habilitacao1 as
HABILITACAO1, c.habilitacao2 as HABILITACAO2, gb.numero as NUMERO, gb.nome as
NOME, t.nome Turma, ta.nome as TURNO, pl.ano as ANO, d.nome as Disciplina,
tdi.nome as TIPODISCIPLINA, d.cargahoraria, p.nome as Professor
FROM turmadisciplina td, turma t, turnoaula ta, disciplina d, professor p,
curso c, periodoletivo pl, tipodisciplina tdi, instituicao i, gradebloco gb
WHERE d.codcurso=c.codigo and c.codinstituicao=i.codigo and
td.codturma=t.codigo and t.codgrdbloco=gb.codigo and t.codturnoaula=ta.codigo
and t.codperletivo=pl.codigo and tdi.coddisciplina=d.codigoand
d.codtipodisciplina=tdi.codigo and td.codprofessor = p.codigo
```

Figura 26 – Select Curso Completo

O processo segue com os componentes *add sequence*, *select values* e *table output* inserindo os dados na base dimensional.

6.8 Fato Matriculado

A tabela fato é composta pelas chaves primárias inseridas pelo processo de ETL das dimensões tempo, acadêmico, curso e a medida: total matriculado foi criado a partir de código javascript. Na Figura 27 são apresentados todos os componentes utilizados.

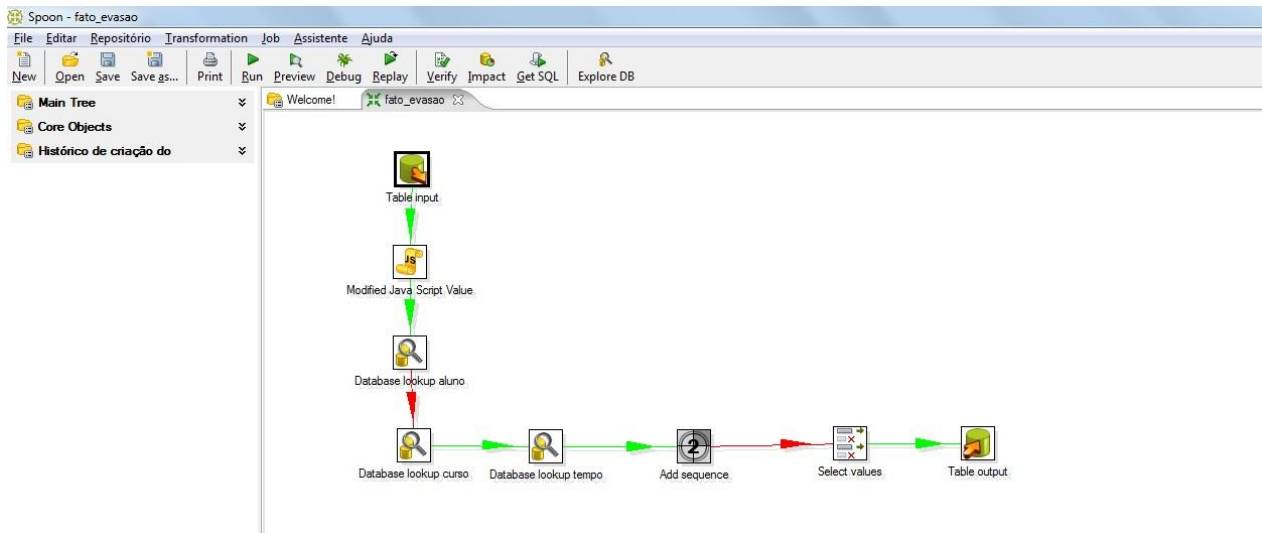


Figura 27 – ETL Fato Matriculado

Para iniciar o processo o componente *table input* foi utilizado e os dados filtrados por meio do comando SELECT como é apresentado na Figura 28.

```
select sm.codaluno, sm.codsituacaoacademica, a.codcurso, (LPAD(extract (month
from a.datamatricula),2,'0')|| '/'|| extract(year from a.datamatricula)) as
DATAMATRICULA
from aluno a, sitacadmovimento sm, situacaoacademica sa, curso c
where sm.codaluno=a.codigo and sm.codsituacaoacademica=sa.codigo and
a.codcurso=c.codigo
```

Figura 28 – Select Fato Matriculado

O próximo componente é um script em *javascript* que por meio do comando abaixo se criou a medida a ser inserida na tabela dimensional como é mostrado na Figura 29.

```

var situacao = CODSITUACAOACADEMICA.getInteger();

var matriculado=0

if (situacao == 1){
    matriculado = 1
}

```

Figura 29 – Script para criação de medida no Fato Matriculado

Em seguida o componente *database lookup* tem a função de associar as chaves primárias das tabelas no fato sendo este processo repetido para todas as dimensões. O componente *add sequence* adiciona uma chave primária para o fato e o *select values* mapeia os campos que estão sendo carregados pelo *table output* realizando a conexão de saída com a base dimensional. Nos dados obtidos no processo de ETL foram encontrados campos vazios referente à dimensão tempo, isso se deve ao cadastro de datas erradas na base acadêmica e devido a este motivo foram retirados todos os registros com este problema que representa apenas 2,3% do total de registros da base sem afetar as consultas neste fato.

6.9 Fato Trancado

Este fato é composto pelas chaves primárias das dimensões: acadêmico, curso, tempo, tempo e tempo final e pelas medidas *total_trancado* e *tempo cursado*, sendo este último calculado pela função *datediff* por meio do comando *SELECT* apresentado na figura 30:

```

SELECT sm.codaluno, sm.codsituacaoacademica, a.codcurso, (LPAD(extract (month
from a.datamatricula),2,'0')|| '/'|| extract(year from a.datamatricula)) as
DATAMATRICULA, (LPAD(extract (month from sm.data),2,'0')|| '/'|| extract(year
from sm.data)) as DATATRANCADO, datediff(month, a.datamatricula, sm.data) as
TEMPOCURSADOMES
FROM aluno a, sitacadmovimento sm, situacaoacademica sa, curso c
WHERE sm.codaluno=a.codigo and sm.codsituacaoacademica=sa.codigo and
a.codcurso=c.codigo

```

Figura 30 – Select Fato Trancado

Após selecionar os dados pelo componente *table input*, em seguida é criada a medida total trancado por meio do componente *javascript*, como é mostrado na Figura 31.

```
var situacao = CODSITUACAOACADEMICA.getInteger();

var trancado=0

if (situacao == 2){
    trancado = 1
}
```

Figura 31 – Script para criação de medida no Fato Trancado

Em seguida são utilizados os componentes, *database lookup*, *add sequence*, *select values e table output* que carrega os dados na base dimensional. Neste fato também ocorreu problemas com a dimensão tempo na qual alguns registros foram deletados que representam 2,5% do total de registros.

6.10 Fato Cancelado

A exemplo do fato anterior o Fato Cancelado é composto pelas chaves primárias das dimensões, acadêmico, curso, tempo e tempo final e pelas medidas: total cancelado e tempo cursado. Na figura 32 é mostrado o SELECT:

```
select sm.codaluno, sm.codsituacaoacademica, a.codcurso, (LPAD(extract (month
from a.datamatricula),2,'0')|| '/')
|| extract(year from a.datamatricula) as DATAMATRICULA, (LPAD(extract (month
from sm.data),2,'0')|| '/')
|| extract(year from sm.data) as DATACANCELADO, datediff(month,
a.datamatricula, sm.data) as TEMPOCURSADOMES
from aluno a, sitacadmovimento sm, situacaoacademica sa, curso c
where sm.codaluno=a.codigo and sm.codsituacaoacademica=sa.codigo and
a.codcurso=c.codigo
```

Figura 32 – Select Fato Cancelado

Em seguida o componente *javascript* para atribuir o número 1 aos acadêmicos que tiveram a matrícula cancelada conforme e mostrado na Figura 33:

```
var situacao = CODSITUACAOACADEMICA.getInteger();

var cancelado=0

if (situacao == 3){
    cancelado = 1
}
```

Figura 33 – Script para criação de medida no Fato Cancelado

Após a criação da medida, foi inseridos os componentes, *database lookup*, *add sequence*, *select values* e *table output*. Foi necessário apagar 2,5% do total de registros, pois também foram encontradas datas inconsistentes.

6.11 Fato Formado

Este fato é constituído pelas chaves primárias das dimensões, acadêmico, curso, tempo e tempo final, sendo as medidas: total formado e tempo cursado. Na Figura 34 é mostrado o SELECT.

```
SELECT sm.codaluno, sm.codsituacaoacademica, a.codcurso, (LPAD(extract (month
from a.datamatricula),2,'0')|| '/'|| extract(year from a.datamatricula)) as
DATAMATRICULA, (LPAD(extract (month from f.dataconclusao),2,'0')|| '/'||
extract(year from f.dataconclusao)) as DATAFORMATURA, datediff(month,
a.datamatricula, f.dataconclusao) as TEMPOCURSADOMES
FROM aluno a, sitacadmovimento sm, situacaoacademica sa, curso c, formatura f
WHERE sm.codaluno=a.codigo and sm.codsituacaoacademica=sa.codigo and
a.codcurso=c.codigo and a.codformatura=f.codigo
```

Figura 34 – Select Fato Formado

Após a seleção dos dados é criado a medida total formado pelo componente *javascript* que é mostrado na Figura 35.

```

var situacao = CODSITUACAOACADEMICA.getInteger();

var formado=0

if (situacao == 4){
    formado = 1
}

```

Figura 35 – Script para criação de medida no Fato Formado

Em seguida foram utilizados os componentes, *database lookup*, *add sequence*, *select values* e *table output*. Este fato também continha registros vazios e foram deletados 0,3% do total de registros.

6.12 Fato Desempenho Acadêmico

O processo de ETL para criação deste fato reúne as chaves primárias das dimensões: acadêmico, curso, curso completo, professor, disciplina, turma e tempo e as medidas: frequência, media, media final, exame, aprovado, reprovado, aprovado com exame, reprovado com exame e reprovado sem exame.

O componente *table input* faz a comunicação com a base acadêmica e os dados são filtrados através do comando SELECT como é mostrado na Figura 36:

```

SELECT tda.codaluno, tda.codturmadisc, td.codprofessor, td.coddisciplina,
td.codigo as CodCursoCompleto, a.codcurso, pl.numperiodo||'-'||pl.ano as
PERIODOLETIVO, tda.media, tda.frequencia, tda.exame, tda.mediafinal
FROM aluno a, turmadiscaluno tda, turmadisciplina td, professor p, disciplina
d, curso c, periodoletivo pl, turma t
WHERE tda.codaluno=a.codigo and tda.codturmadisc=td.codturma and
td.codturma=t.codigo and t.codperletivo=pl.codigo and td.codprofessor=p.codigo
and td.coddisciplina=d.codigo and a.codcurso=c.codigo

```

Figura 36 – Select Fato Desempenho Acadêmico

As medidas aprovado, reprovado, aprovado com exame, reprovado com exame e reprovado sem exame não existem na base acadêmica e, portanto foram criadas para medir o desempenho dos alunos. Para isso foi inserido um script a fim de criar estas medidas, como é apresentado na Figura 37:


```

var mediafinal = MEDIAFINAL.getString();
var media = MEDIA.getString();
var exame = EXAME.getString();
var aprovado=0;
var reprovado=0;
var aprovadoexame=0;
var reprovadosemexame=0;
var reprovadoexame=0;
if (media >= 7){
    aprovado = 1;
}
if (media > 2.9 && media < 7){
    reprovado=1;
    if(exame != null){
        var result = (media+exame)/2;
        if (result >= 5){
            aprovadoexame=1;
        }
        else{
            reprovadoexame=1;
        }
    }
    if(media < 3 || exame == null) {
        reprovadosemexame=1;
    }
}

```

Figura 37 – Script para criação de medidas no Fato Desempenho Acadêmico

O próximo componente é o *database lookup* para encontrar as chaves primárias das tabelas e compará-las com a chave criada no fato. Este processo é repetido para todas as dimensões. Para finalizar o componente *add sequence* adiciona uma chave ao fato, o *select values* mapeia os campos e o *table output* carrega os dados na base dimensional na tabela fato correspondente. Na Figura 38 são mostrados os componentes utilizados:

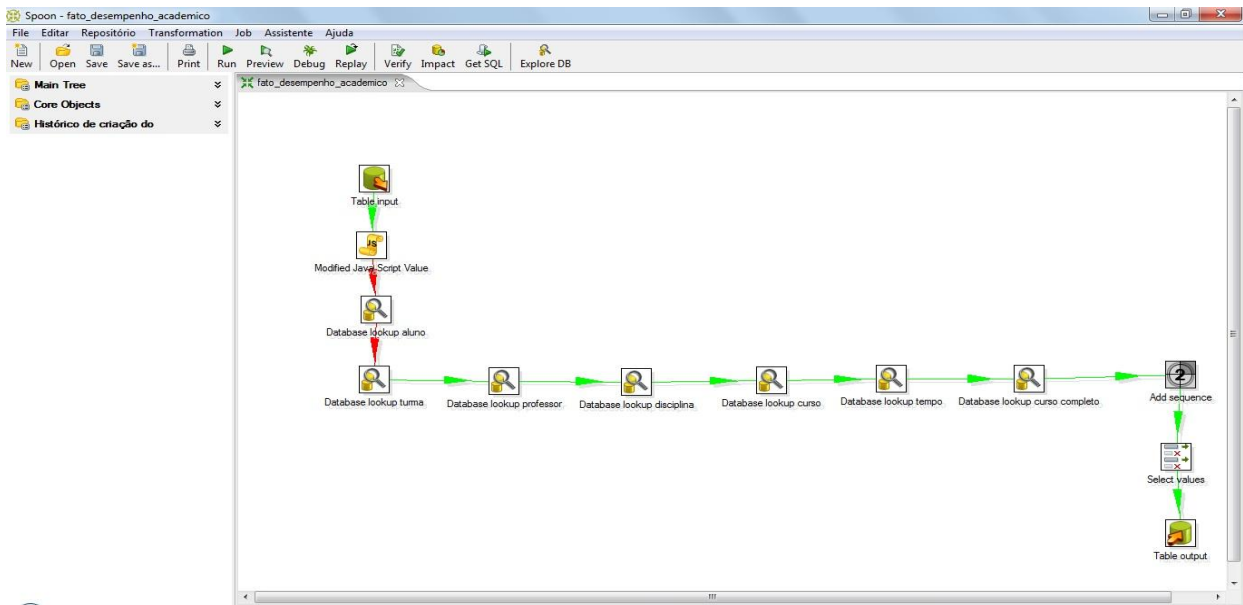


Figura 38 – ETL Desempenho Acadêmico

6.13 Conclusão

A etapa de ETL consistiu em carregar as dimensões e fatos modelados num processo complexo de extração de dados, na qual a filtragem exigiu uma grande atenção para buscar as informações necessárias que às vezes estavam dispersas em várias tabelas precisando reuni-las em conjunto através do comando `SELECT` realizando vários `JOINS`. Durante o processo de busca dos dados foi observado uma consideração sobre a base acadêmica como, por exemplo, a tabela professor não existe um campo referenciando sua titulação, o que poderia ser mais um agregador para realizar análises.

Contudo as transformações foram quase nulas somente sendo necessárias na transposição dos fatos, pois certas medidas foram criadas por serem fundamentais no processo de análise e por não existirem na base acadêmica.

7 Implantação do Data Mart na ferramenta OLAP

Após realizar todo o processo de ETL, a próxima etapa é a criação do *schema* XML para consultas na ferramenta OLAP. Para isso foi utilizado a ferramenta *Schema Workbench* que tem como função facilitar a construção do código XML que será interpretado pelo servidor Mondrian que está integrado com a ferramenta OLAP Brain. Para iniciar o processo é necessário configurar o Workbench com a base dimensional inserindo a *driver class name*, *connection url*, *user name* e *password* como é mostrado na Figura 39.

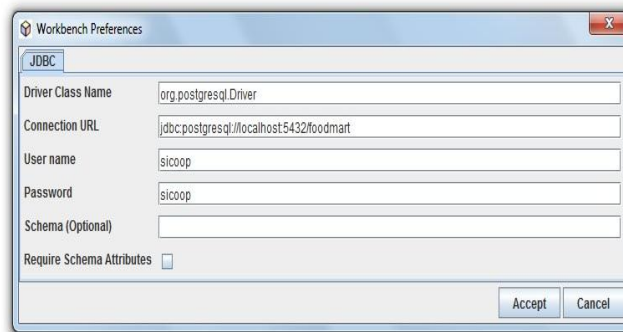
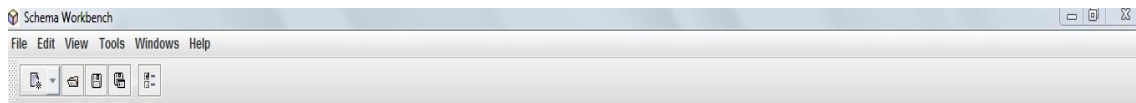


Figura 39 – Conexão com a base dimensional

A partir da conexão com a base dimensional se inicia a construção do XML. Ao clicar com o botão direito do mouse no ícone *schema* é mostrado uma lista de itens na qual é possível inserir cubos, medidas, dimensões, hierarquias e níveis. Na Figura 40 é mostrado o *schema*.

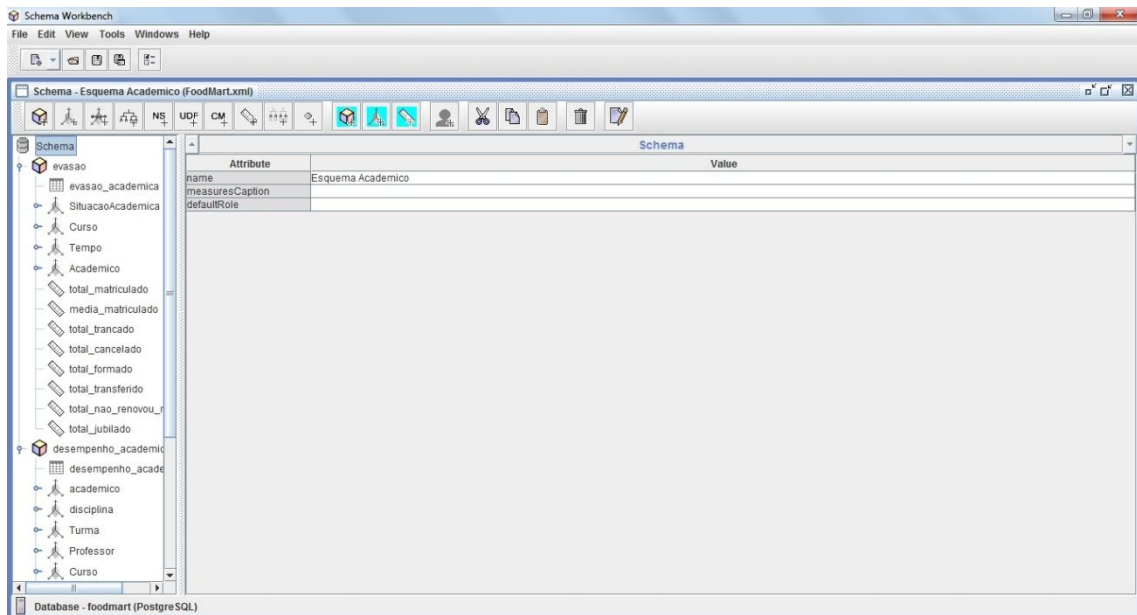


Figura 40 – Schema XML

Então o *schema* é salvo e inserido nos diretórios do servidor mondrian para a ferramenta OLAP “enxergar” o *schema* XML, interpretar e exibir os dados possibilitando realizar as análises.

8 Análise

Após todo o processo de implantação do DM a ferramenta Brain é responsável por exibir os dados organizados em forma de relatórios ou gráficos permitindo utilizar as dimensões e medidas criadas para realização das análises. A primeira atividade no Brain é configurar o diretório do XML, clicando em configurações, fonte de dados, exibirá uma janela na qual é necessário configurar a URL do XMLA para assim selecionar o catálogo criado. Após isso ir em análises, inserir os cubos por meio do botão nova análise. Em seguida são mostradas algumas análises realizadas pelo Brain.

8.1 Formandos por curso

Para esta análise são considerados alunos que se formaram dentre os cursos do Campus Luiz Meneghel, pois a quantidade de alunos é a medida responsável por determinar qual curso tem um maior número de formandos.

Ao filtrar os dados foi determinado o curso de Sistemas de Informação para comparar a evolução do total de formandos durante os anos como é mostrado na Tabela 1.

		Measures
Cursos	Tempo_Final	total_formado
All Curso.Cursoss	All TempoFinal.Tempo_Finals	3.096
AGRONOMIA	All TempoFinal.Tempo_Finals	1.247
CIÊNCIAS BIOLÓGICAS	All TempoFinal.Tempo_Finals	643
ENFERMAGEM	All TempoFinal.Tempo_Finals	484
MEDICINA VETERINÁRIA	All TempoFinal.Tempo_Finals	427
SISTEMAS DE INFORMAÇÃO	All TempoFinal.Tempo_Finals	295
	2005	42
	2006	65
	2007	51
	2008	40
	2009	41
	2010	32
	2011	24

Tabela 1 – Formandos por Curso

Portanto por meio da Tabela 1 é possível verificar que o curso de Agronomia tem o maior número de formandos, isso se deve por ser o curso mais antigo do Campus. No curso de Sistemas de Informação pode-se verificar que houve uma queda acentuada nos últimos anos no número de formandos. Em seguida os dados apresentados em formato gráfico são mostrados no Gráfico 1.

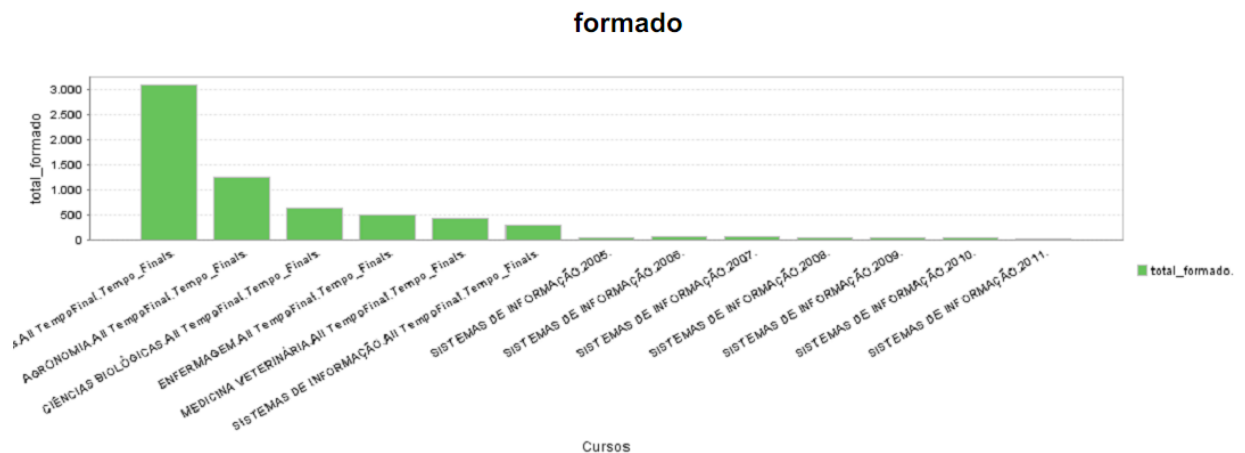


Gráfico 1 – Formandos por Curso

9 CONCLUSÃO

Neste trabalho foi apresentado o conceito de *Business Intelligence* e suas características como uma abordagem necessária para melhor entender a situação do Campus por meio da implantação de *Data Warehouse* e todo o seu processo na qual foi aplicado na base de dados da UENP, mais especificamente na Secretaria Acadêmica do Campus Luiz Meneghel para gerar análises por meio da ferramenta Brain OLAP de forma a ajudar os administradores do Campus na tomada de decisão, dessa forma, atingindo os objetivos especificados neste trabalho.

As maiores dificuldades encontradas foram na etapa de ETL, pois se trata de processo minucioso na busca pelos dados espalhados pela tabelas da base acadêmica e constante verificação da quantidade de dados retornados para possibilitar uma análise mais consistente e de qualidade.

Este trabalho contribuiu para proporcionar uma nova visão do Campus trazendo os dados que estavam apenas armazenados no banco de dados sendo utilizados de forma operacional em informações importantes e úteis como, por exemplo, a evasão de alunos referente aos cursos durante determinado período, nesse tipo de análise o administrador consegue verificar qual curso está com o maior número de ocorrências possibilitando buscar novas soluções e tomar decisões a fim de melhorar e qualidade.

Portanto ao estudar os conceitos envolvidos e analisar seus resultados pode-se observar que o BI é uma opção importante para conhecer melhor o negócio oferecendo uma visão estratégica na busca de um diferencial competitivo. Com relação ao meio acadêmico, neste trabalho pode-se verificar que o BI trouxe uma nova visão do Campus Luiz Meneghel, fato que antes não era possível e o impacto desses resultados pode dar suporte aos gestores para melhorar a qualidade de ensino.

9.1 Trabalhos futuros

Ao finalizar este trabalho implantando a ferramenta Brain OLAP na UENP – Campus Luiz Meneghel e geração de análises com isso é possível sugerir após os estudos realizados os seguintes tópicos para trabalhos futuros:

- Criação de DM em outros dois Campus para se obter uma visão global da Universidade.
- Estudo para utilizar a base dimensional criada aplicando as técnicas de mineração de dados para prever algumas situações.
- Estudo para criação de cargas periódicas na etapa de ETL.
- Criação de novos fatos.
- Estudo para comparação de consultas de dados utilizando a base operacional por meio de SQL e base dimensional por uma ferramenta OLAP analisando as vantagens e desvantagens de utilização.

REFERÊNCIAS

ANTONELLI, Ricardo Adriano. Conhecendo o Business Intelligence (BI): Uma Ferramenta de Auxílio à Tomada de Decisão. **Revista Tecap**, Pato Branco, v. 03, n. 03, p.79-85, 2009. Anual.

BARROS, A. J. S. e LEHFELD, N. A. S. **Fundamentos de Metodologia**: Um Guia para a Iniciação Científica. 2 Ed. São Paulo: Makron Books, 2000.

BRAIN. **A ferramenta Brain**. Disponível em: <http://www.brainbr.com.br/site/index.php?option=com_content&view=article&id=63&Itemid=41>. Acesso em: 22 out. 2011.

CENTENARO, Antonio Cesar et al. **Modelando um Data Warehouse Dimensional a partir de um Modelo Hierárquico**. Disponível em: <<http://inf.unisul.br/~ines/workcomp/cd/pdfs/3030.pdf>>. Acesso em: 17 jun. 2012.

IBM CORPORATION. **DB2 Alphablox**. Disponível em: <<http://www-01.ibm.com/software/data/db2/alphablox/>>. Acesso em: 22 out. 2011.

IBM CORPORATION. **Star Schema Access**. Disponível em: <http://publib.boulder.ibm.com/infocenter/dzichelp/v2r2/index.jsp?topic=%2Fcom.ibm.db2z10.doc.perf%2Fsrc%2Ftpc%2Fdb2z_starschemaaccess.htm>. Acesso em: 22 out. 2011.

INMON, William H. **Buiding the data Warehouse**. Fourth Edition. Indianapolis: Wiley Publising, 2005. 543 p.

INMON, William H. **Como construir o Data Warehouse**. 2ª ed. New York: Editora Campus, 1997.

KIMBALL, Ralph. **Data warehouse toolkit**. São Paulo: Makron Books, 1998.

KIMBALL, Ralph; ROSS, Margy; THORNTHWAITE, Warren; MUNDY, Joy; BECKER, Bob. **The Data Warehouse Lifecycle Toolkit**. Second Edition. Indianapolis: Wiley Publishing, 2008.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e Projeto de Data Warehouse: Uma visão multidimensional**. 5. ed. São Paulo: Érica, 2010.

MENOLLI, André Luis Andrade. **Definição de uma Arquitetura de Data Warehousing para Gestão em Ciência e Tecnologia no Brasil**. 2004. 114 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Estadual de Maringá, Maringá.

MICROSOFT. **Microsoft SQL Server 2008 Reporting Services Report Builder 2.0**. Disponível em: <http://www.microsoft.com/download/en/details.aspx?displaylang=en&id=1932>. Acesso em: 21 out. 2011.

MOREIRA, Eduardo. **Gerência de TI: Modelo Dimensional para Data Warehouse**. Disponível em: http://imasters.com.br/artigo/3836/modelo_dimensional_para_data_warehouse. Acesso em: 02 nov. 2011.

OPENI. **Open Jasper Plug-In**. Disponível em: <http://openi.org/product/>. Acesso em: 22 out. 2011.

ORACLE. **Oracle OLAP**. Disponível em: https://shop.oracle.com/pls/ostore/product?p1=oracleolap&p2=&p3=&p4=&sc=ocom_oracleolap&tz=-2:00. Acesso em: 21 out. 2011.

PENTAHO CORPORATION. **Mondrian**. Disponível em: <http://mondrian.pentaho.org/documentation/>. Acesso em: 09 nov. 2011.

PENTAHO CORPORATION. **Pentaho Analysis**. Disponível em: <http://www.pentaho.com/explore/products/>. Acesso em: 22 out. 2011.

PRASS, Fernando Sarturi; VARGAS, Pablo Tândolo de; FERNANDES, Mikael de Souza. **Utilizando a ferramenta Mondrian para processamento de um Data Warehouse**: WOLAP com ferramenta open-source. Programação + Banco de Dados, São Paulo, n. , p.53-59, 30 jul. 2009.

SERAIN, João Sidemar. **O BI e a Análise do Ambiente**. Disponível em: <http://imasters.com.br/artigo/6988/bi/o_bi_e_a_analise_do_ambiente/>. Acesso em: 09 jun. 2012.