

**FACULDADES INTEGRADAS DE CARATINGA**  
**FACULDADE DE CIÊNCIA DA COMPUTAÇÃO**

**APLICAÇÃO DE FERRAMENTAS DE BUSINESS  
INTELLIGENCE PARA CONSOLIDAÇÃO DE DADOS  
FINANCEIROS DE UMA INSTITUIÇÃO DE ENSINO:  
ESTUDO DE CASO COM A SUÍTE PENTAH0**

**HUDSON SILVA DE SOUZA**

Caratinga

2012

**Hudson Silva de Souza**

**APLICAÇÃO DE FERRAMENTAS DE BUSINESS INTELLIGENCE PARA  
CONSOLIDAÇÃO DE DADOS FINANCEIROS DE UMA INSTITUIÇÃO DE  
ENSINO: ESTUDO DE CASO COM A SUÍTE PENTAHO**

Monografia apresentada à Faculdade de Ciência da Computação das Faculdades Integradas de Caratinga como exigência parcial para obtenção do título de Bacharel em Ciência da Computação, sob orientação do professor Glauber Luiz da Silva Costa.

Caratinga

2012

**Hudson Silva de Souza**

APLICAÇÃO DE FERRAMENTAS DE BUSINESS INTELIGENCE PARA  
CONSOLIDAÇÃO DE DADOS FINANCEIROS DE UMA INSTITUIÇÃO DE  
ENSINO: ESTUDO DE CASO COM A SUÍTE PENTAHO

Monografia submetida à  
Comissão examinadora designada pelo  
Curso de Graduação em Ciência da  
Computação como requisito para  
obtenção do grau de Bacharel.

---

Orientador Prof. Glauber Costa  
Faculdades Integradas de Caratinga

---

Professor Msc. Fabricia Pires Souza Tiola  
Faculdades Integradas de Caratinga

---

Professor Msc. Hebert Luiz Amaral Costa  
Faculdades Integradas de Caratinga

Caratinga, \_\_/\_\_/\_\_

## **DEDICATÓRIA**

Dedico meu trabalho primeiramente a Deus, à minha esposa Viviane, e a meus pais, todos sempre estiveram ao meu lado nos momentos em que mais precisei. Dedico a eles minha vitória.

## **AGRADECIMENTOS**

Agradeço a todos que direta ou indiretamente contribuíram para que eu chegasse até aqui. Em especial agradeço a Deus por tudo o que tem me dado, à minha esposa pelo companheirismo, e aos meus pais, que fizeram tudo que esteve ao alcance deles para que eu atingisse esse objetivo.

Obrigado a todos.

## RESUMO

Este trabalho apresenta a aplicação de *Business Intelligence* na análise financeira de uma instituição de ensino, mostrando a consolidação dos dados de diversos bancos de dados e utilizando esses dados centralizados para criar relatórios que podem ser utilizados em análises financeiras. Nesse trabalho utilizou-se a suíte Pentaho, um conjunto de ferramentas *Open Source* destinada a realizar o processo de *Business Intelligence*, para consolidar os dados extraídos do sistema AdX, o sistema acadêmico da rede de ensino Doctum, dando ao usuário a capacidade de gerar suas análises e auxiliando a tomada de decisões. Os resultados obtidos mostram que *Business Intelligence* é aplicável na rede de ensino Doctum consolidando dados e dando suporte à tomada de decisões, e que a suíte Pentaho contribui para esse processo através de ferramentas bem elaboradas e com grande poder de análise de dados.

**Palavras-chave:** *Business Intelligence*, *Data Warehouse*, OLAP, Painéis Gerenciais, ETL, Modelagem Dimensional, Pentaho.

## ABSTRACT

This paper presents the application of business intelligence in the financial analysis of an educational institution, showing the consolidation of data from various databases and using these centralized data to create reports that can be used in financial analysis. In this paper we used the Pentaho suite, a set of tools designed to make Open Source Business Intelligence process to consolidate the data extracted from AdX system, the system of academic education network Doctum, giving the user the ability to generate their analyzes and assisting decision making. The results show that Business Intelligence is applicable in the rede de ensino Doctum consolidating data and supporting decision making and the Pentaho suite contributes to this process through well-designed tools with great power and data analysis.

**Key-words:** *Business Inteligence, Data Warehouse, OLAP, Dashboards, ETL, Dimensional Modeling, Pentaho.*

## LISTA DE SIGLAS

BI – Business Intelligence (Inteligência Empresarial)

ETL – Extract, Transformation and Load (Extração, transformação e carregamento)

DW – Data Warehouse (Armazém de dados)

OLAP – On-line Analytical Processing (Processamento analítico em tempo real)

DM – Data Mart (repositório de dados)

AdX – Sistema acadêmico da rede de ensino Doctum

SQL – Structured query language (Linguagem de consulta estruturada)

HTML – Hyper Text Markup Language (Linguagem de marcação de hipertexto)

PDF – Portable Document Format (Formato portátil de document)

CSV – Comma-separated values (Valores separados por vírgula)

## LISTA DE ILUSTRAÇÕES

Figura 1: Processo de Data Warehouse (SOFTWARE WIKI, 2012). .....	19
Figura 2: (KIMBALL; ROSS, 2002) Representação da Tabela Fato e Dimensões em um Modelo Dimensional. ....	20
Figura 3: Exemplo de Data Mart – Esquema Estrela (HORITA, 2011).....	22
Figura 4: Exemplo de Data Mart – Esquema Floco de Neve (HORITA, 2011).....	23
Figura 5: Representação do Processo de ETL (RODRIGUES, 2012).....	25
Figura 6: Representação Gráfica de um Cubo OLAP.....	29
Figura 7: Modelo Relacional – Relacionamento entre Tabelas geradoras de informações financeiras do AdX (AdX).....	37
Figura 8: Modelo Dimensional - Esquema Estrela criado para o Data Warehouse .....	39
Figura 9: (Pentaho Kettle) Conexões utilizadas no processo de ETL.....	41
Figura 10: (Pentaho Kettle) Geração da Dimensão Alunos.....	42
Figura 11: (Pentaho Kettle) Geração da Dimensão Matrícula .....	43
Figura 12: (Pentaho Kettle) Geração da Tabela Fato.....	45
Figura 13: (Pentaho Kettle) Encapsulamento de Transformações em um único Job .....	48
Figura 14: (Pentaho Kettle) Encapsulamento de Jobs em um único Job. ....	49
Figura 15: (Pentaho BI Server) Modelo de Dados - Configurando Conexão.....	51
Figura 16: (Pentaho BI Server) Modelo de Dados - Fase Final.....	52
Figura 17: (Pentaho BI Server) Relatório - Selecionando Fonte de Dados e Design .....	53
Figura 18: (Pentaho BI Server) Relatório - Selecionando Campos. ....	54
Figura 19: (Pentaho BI Server) Relatório - Aplicando Filtros.....	55
Figura 20: (Pentaho BI Server) Relatório - Configuração de Página.....	56
Figura 21: (Pentaho BI Server) Relatório de boletos por Filial .....	56
Figura 22: (Pentaho BI Server) Cubo OLAP - Boletos em aberto por Curso.....	57
Figura 23: (Pentaho BI Server) Cubo OLAP - Distribuição por Curso relacionadas em cada Filial .....	58

Figura 24: (Pentaho BI Server) Painel Gerencial – Criação de Layout para exibição. ....	60
Figura 25: (Pentaho BI Server) Painel Gerencial – Criação de Fontes de Dados.....	61
Figura 26: (Pentaho BI Server) Painel Gerencial: Criação de Componentes. ....	62
Figura 27: (Pentaho BI Server) Painel Gerencial – Visualização do Painel Gerencial Pronto. ....	63
Figura 28: (Pentaho BI Server) Relatório de número de Boletos por Curso com quebra por Filial.....	65
Figura 29: (Pentaho BI Server) Relatório de Boletos por Filiais com quebra por Curso.....	66
Figura 30: (Pentaho BI Server) Numero de Boletos em aberto por Curso .....	68
Figura 31: (Pentaho BI Server) Cubo OLAP - Custos x Parcelas.....	69
Figura 32: (Pentaho BI Server) Cubo OLAP - Cursos x Filiais .....	70
Figura 33: (Pentaho BI Server) Cubo OLAP - Cursos e Turmas x Filial.....	70
Figura 34: (Pentaho BI Server) Cubo OLAP – Cursos agrupados por Parcelas x Turmas .....	71
Figura 35: (Pentaho BI Server) Cubos OLAP - Gráfico gerado na interface geradora de Cubo OLAP.....	72
Figura 36: (Pentaho BI Server) Painel Gerencial - Inadimplência Geral .....	73
Figura 37: (Pentaho BI Server) Painel Gerencial - Inadimplência por Filial.....	74
Figura 38: (Pentaho BI Server) Painéis Gerenciais - Distribuição por Turma.....	75

## SUMÁRIO

1	INTRODUÇÃO.....	13
2	REFERENCIAL TEÓRICO.....	17
2.1	<i>Data Warehouse</i> .....	17
2.1.1	Modelo Dimensional.....	19
2.1.2	<i>Data Marts</i> .....	22
2.1.3	ETL - Processo de Povoamento do <i>Data Warehouse</i> .....	24
2.2	Business Intelligence.....	27
2.3	<i>On-line Analytical Processing</i> - OLAP.....	28
2.4	Ferramentas DE BUSINESS INTELLIGENCE.....	29
2.4.1	A suíte Pentaho.....	30
2.4.2	Sistema Acadêmico ADX.....	33
3	METODOLOGIA.....	35
3.1	Análise do Sistema ADX.....	35
3.2	Criando o <i>Data Warehouse</i> .....	38
3.3	Povoando o <i>Data Warehouse</i> .....	40
3.3.1	Dimensão Alunos.....	41
3.3.2	Dimensão Matrícula.....	42
3.3.3	Tabela Fato.....	45
3.3.4	Criando Jobs (Trabalhos) e Encapsulando as Transformações.....	47
3.3.5	Criando o Arquivo Principal.....	48
3.4	Modelando os dados do <i>Data Warehouse</i> .....	50
3.4.1	Utilizando a Informação Gerada.....	52
4	RESULTADOS OBTIDOS.....	64
4.1	Relatórios.....	64
4.2	Cubo OLAP.....	66
4.3	Painéis Gerenciais.....	72
5	CONCLUSÃO.....	76
6	TRABALHOS FUTUROS.....	78

REFERÊNCIAS .....	79
-------------------	----

## 1 INTRODUÇÃO

O crescimento natural das empresas nas últimas décadas e o conseqüente acréscimo da complexidade em seu gerenciamento fez com que as empresas se adequassem à nova realidade vivida, onde o controle é indispensável, e dessa forma trocaram o controle manual pelos sistemas de informação. Com a implantação de sistemas de informação, as empresas tornaram vários processos mais rápidos e eficazes.

Em alguns casos, em que as empresas buscavam soluções viáveis e ao mesmo tempo com baixo custo, houve integração entre vários sistemas para atender a todos os setores e atividades das mesmas, e algumas empresas passaram a trabalhar com mais de um sistema. Dessa forma, os dados podiam estar armazenados em bancos de dados separados e até mesmo bancos de dados de tecnologias diferentes.

Com o passar do tempo, as empresas possuíam um grande volume de dados provenientes da alimentação diária de seus sistemas. As empresas passaram a procurar formas de analisar esse grande volume de dados, transformando-os em informações úteis para tomar decisões (OLIVEIRA, 2006).

Os diferentes sistemas operacionais em que os sistemas são hospedados, as diferentes linguagens de programação em que eles são desenvolvidos, seus bancos de dados separados ou mesmo de tecnologia diferente, fazem com que o custo para se consolidar os dados seja grande, ou até inviável.

Também se pode encontrar informações úteis em planilhas eletrônicas e arquivos de texto espalhados em diferentes computadores dentro de uma empresa. Outro fator que interfere na consolidação dos dados de sistemas diferentes é a forma como os dados são armazenados, pois sistemas podem tratar alguns dados semelhantes de forma diferente, como por exemplo, o sexo de uma pessoa. Um sistema pode armazenar masculino e feminino, outro simplesmente “m” ou “f” ou até mesmo “masc” ou “fem”. Isso dificulta a junção desses dados, pois os padrões não podem ser agrupados facilmente.

A busca de informações consolidadas também ocorre nas instituições de ensino, onde os sistemas transacionais armazenam um grande volume de dados

não só acadêmicos, mas também de dados financeiros dos alunos. As instituições de ensino precisam analisar esses dados financeiros para ter controle de sua situação financeira. Essa informação financeira pode impactar em diversos aspectos, como no ajuste do valor da mensalidade mensal do aluno, no nível de agressividade da estratégia de cobrança de inadimplentes e até mesmo no valor pago aos professores.

As técnicas de *Business Intelligence* podem ajudar a resolver os problemas de análise consolidada de dados de sistemas diferentes, de vários bancos de dados de um mesmo sistema ou arquiteturas diferentes de bancos de dados. As ferramentas de BI permitem extrair dados de diversos locais, como bancos de dados, arquivos de texto, planilhas eletrônicas, padronizar os dados e armazená-los em um único local, seguindo um formato único e propiciando a acesso à leitura dos mesmos. Sobre os dados armazenados emprega-se ferramentas de análise, permitindo criar mecanismos que possibilitem ao usuário final elaborar suas próprias análises e obter seus resultados sem a necessidade de desenvolvimento de novos relatórios.

Com o intuito de demonstrar a viabilidade do uso de ferramentas de Business Intelligence como suporte à análise de situação financeira de uma instituição de ensino, nesse trabalho aplicou-se a suíte Pentaho para analisar os dados financeiros do sistema AdX na rede de ensino Doctum.

Para tanto foram estudadas as regras de negócio que integram as operações financeiras da rede de ensino Doctum. Foram levantados os processos que compõem a emissão de boletos, suas baixas e a contabilização dos números de pagamento e boletos em aberto. Entender o processo e todas as suas etapas é indispensável para que se possa extrair os dados da forma correta. Além de conhecer o processo é necessário conhecer o sistema em que os dados são armazenados.

O sistema responsável pela emissão e controle dos boletos da rede de ensino Doctum é o sistema AdX. Ele é utilizado desde o início da instituição e foi desenvolvido para gerenciar as filiais individualmente. Assim cada instância do sistema gerencia uma unidade de negócio de uma filial, e possui um banco de dados. Uma unidade de negócio é um dos seguimentos de ensino que a instituição possui: colégio, graduação, pós graduação ou mestrado. O AdX

contém além do gerenciamento acadêmico, o gerenciamento financeiro da instituição. No AdX são emitidos os boletos mensalmente, e ele garante o controle dos pagamentos dos mesmos.

Os bancos de dados do AdX são padronizados e altamente normalizados, garantindo a eficácia da operação diária de controle acadêmico e financeiro. Cada instância possui módulo de relatórios, para realizar a análise individual nos bancos de dados. Esses relatórios são desenvolvidos à medida que são solicitados, de acordo com a necessidade do usuário. Não há módulo de relatórios unificados de mais de uma instância do AdX. A consolidação de resultados é feita criando-se manualmente relatórios através dos relatórios individuais emitidos em cada instância.

Para se conseguir análises de forma mais rápida e de forma consolidada nesse trabalho utilizou-se *Business Intelligence*. As ferramentas de *Business Intelligence* tornam a análise mais eficiente e intuitiva, ajudando a tomada de decisões empresariais, pois geralmente utilizam como fonte de dados os *Data Warehouses*. (SCHEPS, 2008).

Nesse trabalho utilizou-se dezenove bancos de dados do AdX, entre colégio, graduação e pós graduação de diversas filiais. Essas eram as instâncias do AdX criadas até o início desse trabalho e por isso foram utilizadas.

A utilização de *Data Warehouse* torna o processo de análise de informações mais eficiente, pois sua estrutura faz com que o acesso aos dados seja centralizado em um único local, e a forma de armazenamento prioriza a velocidade de acesso aos dados (MACHADO, 2007).

Para extrair e armazenar os dados do AdX no *Data Warehouse* utilizou-se rotinas automatizadas, através a ferramenta Kettle da suíte Pentaho. Com essa ferramenta é possível se conectar a vários tipos de fontes de dados, sejam bancos de dados, planilhas eletrônicas ou arquivos de texto e após a extração dos dados dessas fontes tratar os dados para formatá-los e inseri-los em um único *Data Warehouse* com um formato que propicie a análise dos mesmos.

Após a análise do AdX criou-se o modelo dimensional para armazenar todos os dados extraídos dos bancos de dados do AdX. Esse modelo armazena os dados de todos os bancos de dados do AdX em um único banco de dados, um *Data Warehouse*, tornando o acesso aos dados centralizado.

Com os dados consolidados em um único local utilizou-se o Pentaho BI Server, também da suíte Pentaho para mostrar as análises que podem ser obtidas com ele.

Esse trabalho foi organizado da seguinte maneira: no capítulo 2 serão apresentados conceitos de vários autores sobre *Business Intelligence* e *Data Warehouse*. Também será apresentando um resumo das ferramentas utilizadas nesse trabalho. No capítulo 3 será exibida a metodologia utilizada para a criação de um *Data Warehouse*, para a criação de um modelo de dados baseado nesse *Data Warehouse* e a construção de um relatório, cubos e painéis gerenciais. No capítulo 4 serão exibidos os resultados obtidos, demonstrando as ferramentas de análise de dados. O capítulo 5 apresenta as conclusões obtidas e o capítulo 6 fala sobre trabalhos futuros que podem ser realizados com base nesse trabalho.

## 2 REFERENCIAL TEÓRICO

A utilização de técnicas de *Business Intelligence* tem-se tornado uma alternativa viável para a análise de informações gerenciais. Elas facilitam o trabalho, com funcionalidades que vão além de simples relatórios gerenciais.

A ferramenta de ETL dá suporte à criação de um *Data Warehouse*, extraindo dados, tratando-os, enriquecendo-os com informações úteis e armazenando de maneira a permitir que sejam consultados rapidamente.

Uma ferramenta de análise de dados transforma os dados armazenados no *Data Warehouse* em informações úteis, através de análises bem elaboradas. Para que essas técnicas funcionem é necessário que os dados estejam armazenados em um único local, e modelados de forma a maximizar o desempenho de consultas (MACHADO, 2007).

Nesse capítulo serão apresentadas ferramentas, técnicas e conceitos que norteiam as técnicas de *Business Intelligence*, como *Data Warehouse*, *Data Marts*, granularidade de dados, ETL e cubos OLAP.

### 2.1 DATA WAREHOUSE

Com o crescente desenvolvimento das empresas e conseqüentemente de seus sistemas, é comum se encontrar várias organizações com diversas filiais, onde cada uma pode possuir diferentes estruturas de sistemas. Também encontra-se dentro de uma mesma organização vários sistemas, com seus bancos de dados com estruturas e arquiteturas completamente diferentes. Além disso, também podem ser encontrados informações armazenadas em planilhas eletrônicas, arquivos de texto ou mesmo documentos Web.

O termo *Data Warehouse* é utilizado para definir o conjunto de informações coletadas de uma ou diversas fontes, tratadas e armazenadas em um único local, seguindo determinado esquema. Dessa forma, os dados armazenados em um

único local permitem acesso de forma rápida a dados históricos (SILBERSCHATZ et al., 2006).

Essas fontes de dados podem possuir esquemas diferentes ou mesmo modelos de dados distintos. Para fazer uma análise desses dados consolidados, é necessário que eles passem por tratamento e sejam armazenados em um único local. Com isso, esses dados tratados podem não ser apenas uma cópia dos dados das fontes (MACHADO, 2007).

*Data Warehouse* não pode ser tratado como apenas um armazém de dados, mas sim, como toda a tecnologia envolvida no processo de armazenamento desses dados. Desde sua extração, até o armazenamento, os dados passam por várias etapas necessárias para que retratem a realidade evidenciada (SCHEPS, 2008). O processo de fazer *Data Warehouse* pode ser considerado como toda a coleção de tecnologia destinada a ajudar um profissional qualificado a tomar decisões empresariais. Uma tecnologia destinada a apoiar a tomada de decisões empresariais (ELMARI; NAVATHE, 2005).

Pode-se definir *Data Warehouse* como um banco de dados utilizado para armazenar dados de um ou mais sistemas, que depois de extraídos e tratados, são estruturados de forma a permitir consultas de forma eficiente (MACHADO, 2007). A Figura 1 ilustra o processo de *Data Warehouse*, onde pode-se ver a extração realizada de diversas arquiteturas de armazenamento, transformados e armazenados no *Data Warehouse*, em categorias denominadas *Data Marts*, para que possam gerar cubos e modelos de relatórios para o usuário final.

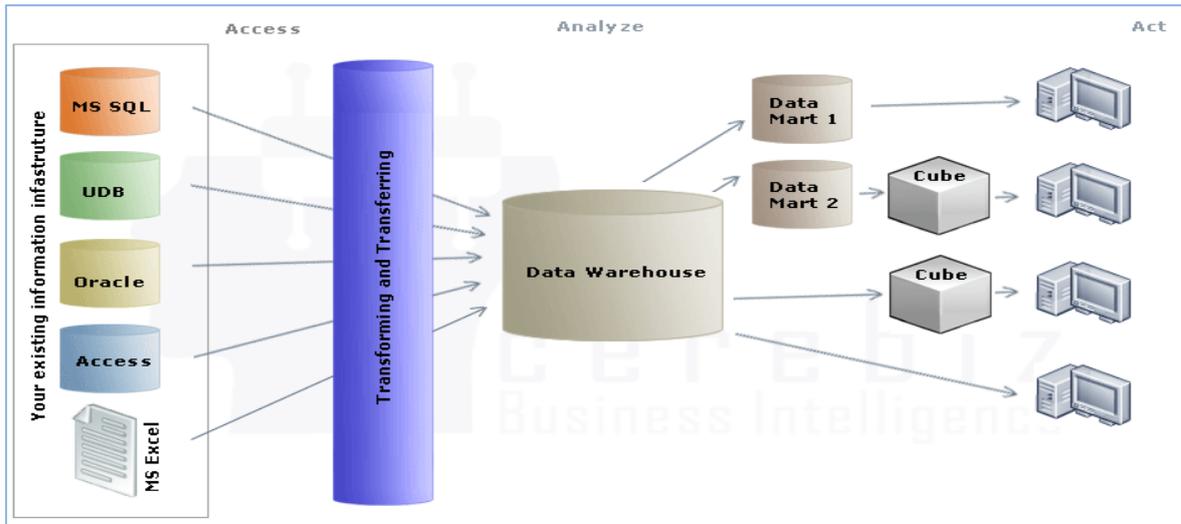


Figura 1: Processo de Data Warehouse (SOFTWARE WIKI, 2012).

Os *Data Warehouse* são não voláteis, o que significa que os dados ficam seguros, pois nele são realizadas apenas duas operações: inserção e leitura de dados. É importante ressaltar que a inserção é realizada em blocos, após os dados serem filtrados dos diversos sistemas e transformados. Isso garante um melhor desempenho do sistema, pois não há operações constantes de atualização de dados (MACHADO, 2007).

Para entender melhor o processo de *Data Warehouse*, a seguir serão mostrados conceitos de *Data Marts*, Modelo Dimensional e processo de ETL. Eles são a base para o entendimento do processo de *Data Warehouse*.

### 2.1.1 Modelo Dimensional

A modelagem dimensional ou multidimensional é uma técnica de projeto lógico utilizada na construção do *Data Warehouse*, diferente da modelagem entidade-relacionamento. Ela possui esse nome por ser formada por tabelas denominadas dimensões ligadas á uma tabela principal, a tabela fato (KIMBALL; ROSS, 2002).

A modelagem dimensional tem por finalidade facilitar a visualização das informações, como se fosse um cubo que pudesse ser observado de vários ângulos. Com essa forma de visualização o modelo multidimensional permite que esse cubo seja fatiado e se trabalhe com um dessas partes (MOREIRA, 2006).

Apesar de ser representado por um cubo, o modelo dimensional é mais simples de entender que a modelagem relacional. Normalmente ele utiliza o *Star Skema*, esquema estrela, para representação dos dados. Também pode ser utilizado o *SnowFlake Skema*, ou esquema floco de neves.

Na Figura 2 pode-se observar um modelo dimensional utilizando o esquema estrela. Nela está representada a tabela fato vendas diárias, a dimensão data e a dimensão produtos. As dimensões estão ligadas diretamente à tabela fato, facilitando a modelagem dos dados.

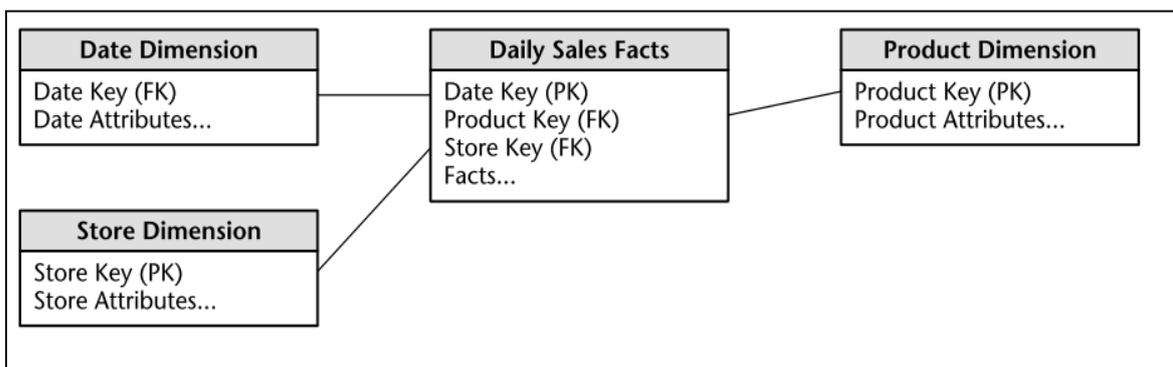


Figura 2: (KIMBALL; ROSS, 2002) Representação da Tabela Fato e Dimensões em um Modelo Dimensional.

A modelagem dimensional é formada por três elementos básicos, sendo eles: os fatos, as dimensões e as medidas.

A tabela de fatos é composta por valores que podem ser medidos e por várias chaves estrangeiras que se relacionam com as outras tabelas do modelo. Essa tabela possui dados numéricos como valores, quantidades e datas. (FREITAS, 2010).

Para MACHADO (2007) cada registro ou fato, daí o nome, representa um item, transação ou evento de negócio. É tudo aquilo que representa a evolução dos negócios diariamente na empresa. Sua característica principal são os dados numéricos.

A tabela de fatos costuma ser a maior do esquema, por conter todas as transações do negócio em questão. Ela se relaciona com as dimensões, que são tabelas menores com as diversas visões do negócio (MOREIRA, 2006).

Para ilustrar uma tabela fato pode-se citar uma tabela de registro de vendas de uma loja. Ela conterá dados de valores vendidos, número de vendas, valor de compra do produto, data de compra e venda, entre outros dados que ajudam a medir o desempenho do negócio.

Para MACHADO (2007) dimensões são elementos participantes do fato, um assunto do negócio. Uma dimensão é uma das possíveis formas de visualização dos dados. O modelo dimensional pode ser entendido como um cubo, e as dimensões são as arestas dessa cubo. Se esse cubo for girado uma nova visualização é revelada.

Uma tabela dimensão contém descrições contextuais sob uma perspectiva do negócio. Seus atributos descrevem detalhadamente os elementos que fazem parte da análise sob o ponto de vista dessa dimensão (MOREIRA, 2006).

Segundo FREITAS (2010), ao se olhar as dimensões do ponto de vista de consultas SQL, seus atributos são as fontes para os cabeçalhos de colunas do conjunto de resposta da consulta.

Para se analisar os fatos sob o ponto de vista de determinada dimensão, deve-se obter resultados para medir determinado assunto. As medidas são campos da tabela fato utilizados como resultado da análise solicitada.

As medidas nada mais são do que números conceitualizados dentro de um contexto. Valores de compra, vendas, contadores de registros são exemplos de medidas. Elas podem ser agrupadas através de funções de soma, contagem média e outras, utilizadas para medir determinada dimensão.

Para MACHADO (2007) a medida é determinada pela combinação das dimensões que participam de um fato. Para ANZANELLO (2008) a medida é uma dimensão especial, que é utilizada para realizar comparações. Nessa dimensão estão incluídas métricas como custos, lucros ou taxas.

### 2.1.2 Data Marts

O *Data Warehouse* pode ser visto como um grande armazém de dados que integra informações relevantes para a empresa, e que é utilizado para tomada de decisões. Esse grande armazém de dados pode e deve ser dividido por categorias, por assuntos.

A essa divisão de assuntos dentro de um *Data Warehouse*, dá-se o nome de *Data Mart*. Os assuntos são divisões dentro de um *Data Warehouse* separando, por exemplo, informações de diferentes departamentos SCHEPS (2008). Assim tem-se um *Data Mart* de vendas, um de estoque, e outro de compras dentro do mesmo *Data Warehouse*.

Para MACHADO (2007) cada *Data Mart* normalmente é composto por um modelo dimensional formado por um esquema estrela, composto por mais de uma tabela conforme Figura 3. Dessa forma, se tem relacionamentos simples, o que melhora a velocidade da busca pela informação através de consultas que fazem junções simples entre essas tabelas.

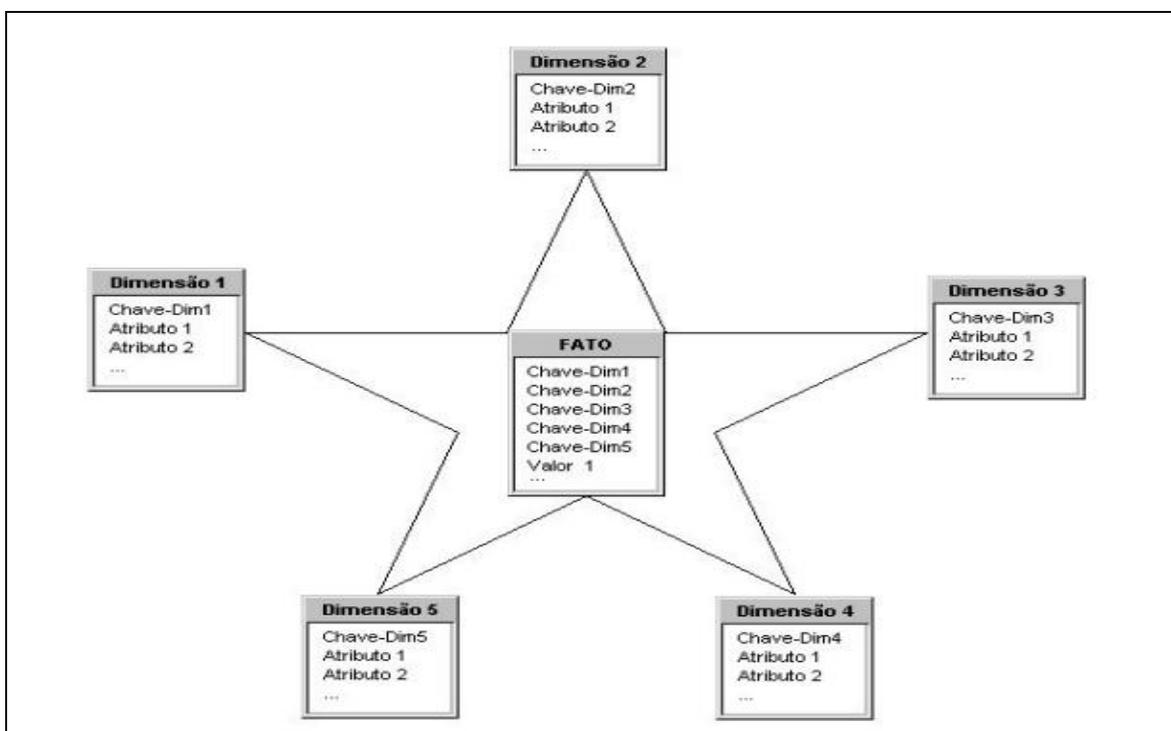


Figura 3: Exemplo de Data Mart – Esquema Estrela (HORITA, 2011).

Além do esquema estrela o *Data Mart* pode ser construído através do modelo floco de neve, conforme ilustrado na Figura 4. Esse modelo é menos utilizado, por conter certa complexidade, diminuindo o desempenho nas consultas (MACHADO, 2007). Nesse modelo algumas tabelas dimensão menores são ligadas a uma dimensão e não diretamente à tabela fato. Isso aumenta o tempo de resposta das pesquisas, e não deve ser utilizada quando se lidar com grande volume de dados (HORITA, 2011).

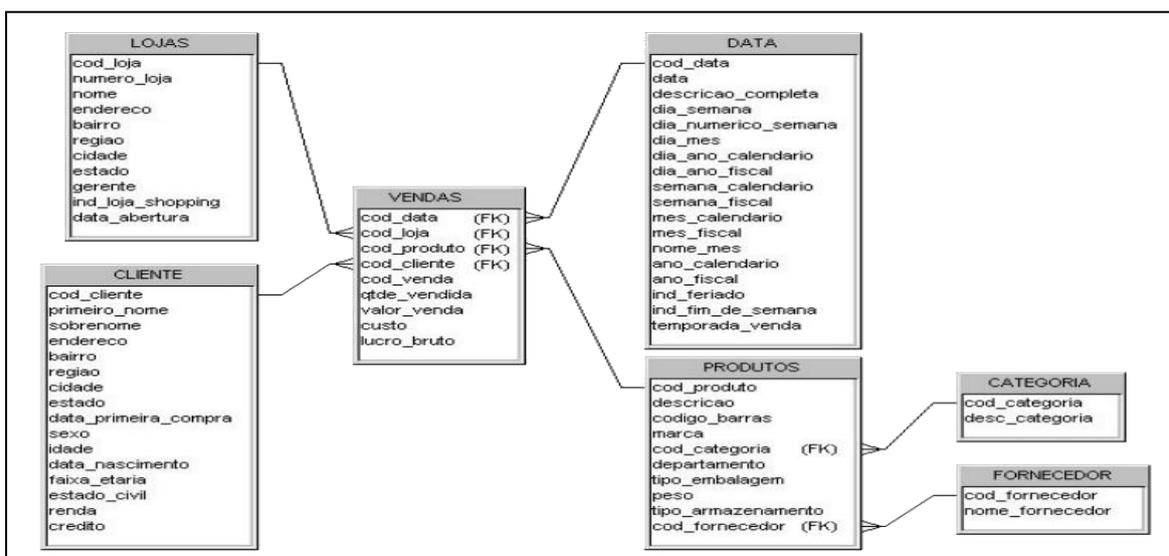


Figura 4: Exemplo de Data Mart – Esquema Floco de Neve (HORITA, 2011).

Para se definir o detalhamento máximo que se deseja em um *Data Mart* deve-se pensar em um fator fundamental para o sucesso do processo de *Data Warehouse*: a granularidade de dados.

Granularidade de dados é um requisito crítico no projeto de *Data Warehouse* e deve ser dada a devida atenção a ele. Isso não quer dizer que quanto maior o detalhamento de dados, melhor será o projeto em questão. Quer dizer que se deve saber previamente o nível de detalhamento que se espera ter em determinado projeto (MACHADO, 2007).

MACHADO (2007) diz que quando fala-se em granularidade de dados, deve-se levar em consideração o nível de detalhamento que se quer ter no projeto. Em alguns casos, um grande detalhamento pode não ser necessário, ou mesmo tornar todo o projeto inviável, devido ao grande número de dados.

Segundo CRAMER (2006), o desempenho das consultas no *Data Warehouse* está diretamente ligado ao nível de granularidade utilizada. Ou seja, quanto maior a granularidade, melhor o desempenho. Porém, uma baixa granularidade permite uma melhor análise no nível mais baixo da informação.

Um *Data Mart* é projetado para se ver determinada informação em grandes períodos de tempo. Dessa forma, um alto nível de granularidade pode fazer com que o acesso aos dados seja mais demorado do que o desejado. Da mesma forma, determinado assunto pode requerer um nível maior de detalhamento, e por isso deve ser tratado com mais cuidado quando se determina a granularidade a ser utilizada. Assim a granularidade depende do propósito a que o projeto irá servir (MACHADO, 2007).

A granularidade diz respeito à forma como os dados estarão agrupados. Caso se queira ver detalhes das transações, deve-se utilizar uma baixa granularidade e dessa forma armazenar os dados detalhados. Se não for requerido muito detalhamento os dados podem ser agrupados e armazenados com maior granularidade. Assim serão armazenados menos registros.

### **2.1.3 ETL - Processo de Povoamento do *Data Warehouse***

Povoar um *Data Warehouse* significa inserir em suas tabelas apenas dados que serão utilizados na análise da organização sob algum ponto de vista. Essa etapa é considerada uma tarefa crítica, pois todas as demais etapas dependem dessa.

As ferramentas de extração, transformação e carregamento de dados (Extraction, Transformation and Load – *ETL*) são úteis e indispensáveis para o povoamento do *Data Warehouse*. Geralmente possuem uma interface gráfica que tornam a tarefa mais agradável. Porém deve-se tomar o cuidado de que ela seja ampla e dotada de grande capacidade, e que propicie acesso rápido às informações desejadas (OLIVEIRA, 2006).

Na Figura 5 pode-se ver o esquema do processo de ETL, entre os bancos de dados transacionais e o *Data Warehouse*. O processo de ETL realiza a ligação entre os bancos de dados dos sistemas transacionais e a *Data Warehouse*, extraindo os dados desses sistemas, transformando os dados, acrescentando informações úteis e armazenando no *DW*.

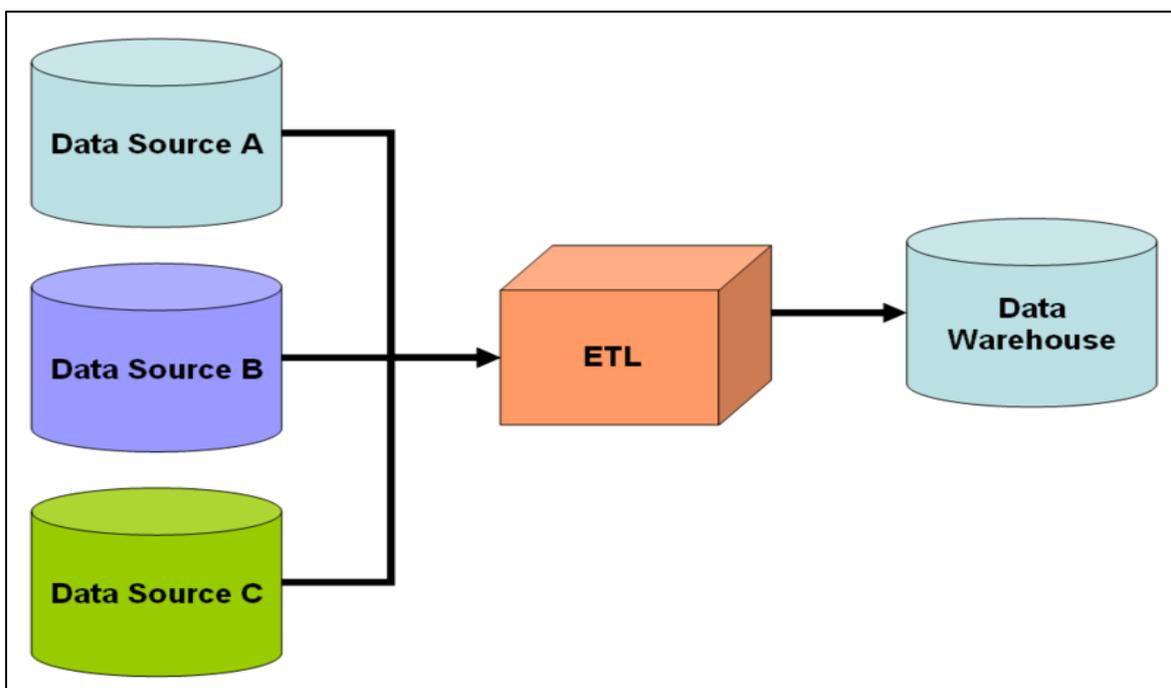


Figura 5: Representação do Processo de ETL (RODRIGUES, 2012).

Para FUJIWARA (2006) o povoamento do *Data Warehouse* é a fase que demanda mais cuidado do gerente de projetos, pois é exigido um tato apurado do gerente, inclusive pelo tratamento com outras áreas da tecnologia.

MACHADO (2007) salienta que a extração, organização e integração dos dados devem ser realizadas para garantir consistência e integridade das informações. Somente assim será possível construir uma base de dados de alta qualidade e confiabilidade.

### 2.1.3.1 Extração de Dados

A extração dos dados deve ser realizada de forma cuidadosa, pois os dados extraídos devem refletir a realidade da informação requerida. As extrações geralmente demandam operações complexas, alteração no formato dos dados, e criação de novas informações, através de funções simples ou complexas.

Existem ferramentas de ETL que fazem conexão com diversos bancos de dados, extraindo os dados já no formato requerido. Dessa forma é possível transferir o processamento para essas ferramentas, aliviando os sistemas transacionais (FUJIWARA, 2006).

A extração requer dos profissionais envolvidos um razoável conhecimento das bases das quais as informações serão extraídas e também armazenadas. Eles devem conhecer os bancos de dados mapeados como fonte de alimentação, e conhecer bem o modelo multidimensional utilizado para conseguir fazer a ligação entre eles (MACHADO, 2007).

### 2.1.3.2 Transformação de Dados

A transformação é a etapa responsável por verificar registro a registro a informação levantada, identificando padrões e garantindo a obrigatoriedade da informação.

Essa etapa verifica padrões, unificando informações não padronizadas, criando padrões e novos campos através dos já existentes. Um exemplo seria sexo representado por “m” ou “f” em um sistema e “masc” e “fem” em outro. Essa etapa é responsável por criar um padrão e aplicá-lo a todos os dados recebidos (FUJIWARA, 2006).

(OLIVEIRA, 2006) acrescenta que a transformação se torna eficiente quando executada antes da carga dos dados no *Data Warehouse*. Ao se evitar as transformações em tempo de execução das consultas executadas pelos usuários, a análise será mais rápida.

### 2.1.3.3 Carregamento de Dados

Após extrair e transformar os dados, esses devem ser inseridos no *Data Warehouse*. Esses dados podem estar em bancos de dados, planilhas eletrônicas, ou mesmo arquivos de texto. A ferramenta utilizada no carregamento de dados geralmente é a mesma responsável pelo processo anterior. De qualquer forma ela deve auxiliar no controle de erros e falhas.

O carregamento de dados pode ser realizado de duas formas: incremental e total. A forma incremental atualizará apenas a tabela fato, não se preocupando com as dimensões. Já a forma total tomará o cuidado de atualizar todas as dimensões (OLIVEIRA, 2006).

## 2.2 BUSINESS INTELLIGENCE

Entende-se por Inteligência de Negócios (*Business Intelligence - BI*), qualquer atividade, ferramenta ou processo utilizado para obter a melhor informação possível dando o suporte ao processo de tomada de decisões.

*Business Intelligence - BI*, que pode ser traduzido como inteligência de negócios, é uma classe de tecnologias e ferramentas utilizadas para atender o nível gerencial das empresas (SCHEPS, 2008).

Para se realizar *Business Intelligence* é necessário que os dados estejam armazenados de forma que propicie um fácil acesso aos mesmos. Por isso não se pode falar em BI sem se falar de *Data Warehouse* (SCHEPS, 2008).

Para (COELHO *et al.*, 2011), as técnicas de *Business Intelligence* englobam recursos como *ETL*, *Data Warehouse*, *Data Marts*, *Data Mining* e OLAP, que juntos dão suporte à tomada de decisões, com informações atualizadas.

### 2.3 ON-LINE ANALYTICAL PROCESSING - OLAP

O processamento analítico on-line (*On-line analytical Processing*, OLAP) pode ser entendido com uma ferramenta capaz de realizar análises de dados com uma visão dimensional do negócio, comparando-os por diversos ângulos (FORTULAN, 2005).

ANZANELLO (2008) diz que OLAP supre os anseios de *Business Intelligence*, utilizando para esse fim os dados armazenados no *Data Warehouse*. Segundo ele, OLAP é uma tecnologia utilizada para visualizar os dados agregados, diferentemente dos dados relacionais, pois a finalidade dessa análise é apoiar a tomada de decisões.

Ainda segundo FORTULAN (2005) as ferramentas OLAP possuem algumas características que as definem:

- Permitem a visão multidimensional dos dados;
- Realizam cálculos complexos;
- Fazem previsões e análises de tendências;
- Constroem cenários a partir de suposições e manipulam dados através de diferentes dimensões.

A tecnologia OLAP permite uma visão multidimensional dos dados, que podem ser visualizados por diversos ângulos, dando assim uma análise rápida e dinâmica dos dados (SOUZA, 2003). As bases de dados OLAP são divididas em cubos, dando uma noção de que de qualquer ângulo que se visualizar o cubo será apresentado uma análise diferente.

MACHADO (2007) salienta que OLAP é um conjunto de ferramentas que permitem explorar os dados do *Data Warehouse*, através de análise multidimensional, representando os dados através de dimensões, e não tabelas. É possível ter uma visão dos dados do *Data Warehouse*, através da combinação das dimensões.

Na Figura 6 pode-se ver a representação de um cubo OLAP. Essa figura representa a forma de visualização propiciada pela visão multidimensional. Para se mudar o foco da análise realizada sob determinado modelo seria como se o

cubo fosse girado e novos dados seriam revelados, sem necessidade de gerar outros relatórios.

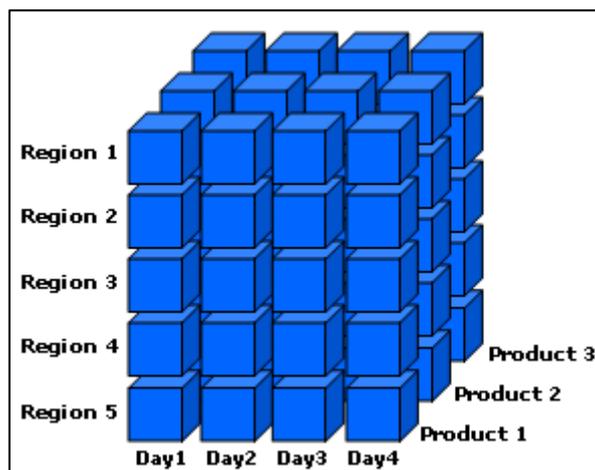


Figura 6: Representação Gráfica de um Cubo OLAP.

A análise OLAP resulta em uma informação que pode determinar o comportamento de determinadas variáveis no decorrer do tempo, permitindo descoberta de tendências e cenários, transformando os dados do *Data Warehouse* em informações estratégicas (MACHADO, 2007).

Segundo SCHEPS (2008), as ferramentas OLAP dão aos sistemas de *Business Intelligence* a capacidade de olhar para os dados de uma forma nova. A potência computacional adicionada às ferramentas inovadoras criou esse novo paradigma.

CRAMER (2006) diz que os servidores OLAP processam e atendem requisições feitas ao *Data Warehouse*, envolvendo cálculos, consolidações de dados e recuperação de informação multidimensional.

## 2.4 FERRAMENTAS DE BUSINESS INTELLIGENCE

Existem no mercado várias ferramentas para *Business Intelligence*. Cada uma possui suas particularidades e foram desenvolvidas para ser utilizadas em determinadas situações.

A Pentaho Corporation se destaca no cenário mundial por possuir uma suíte de ferramentas *Open Source* voltada para *Business Intelligence*. Nesse trabalho foram utilizadas duas de suas ferramentas: O Pentaho Kettle e o Pentaho BI Server.

Foi escolhida a suíte Pentaho por conter ferramentas Open Source com todos os recursos necessários para se realizar *BI*, além de ser a mais utilizada em todo o mundo. Essas ferramentas serão descritas abaixo. Também será apresentada uma descrição do sistema acadêmico AdX, por ser utilizado como fonte de dados para esse trabalho.

#### 2.4.1 A suíte Pentaho

A suíte Pentaho é uma solução *Open Source* que atende todas as etapas para a construção de *Business Intelligence* e por isso diferente das ferramentas de *Business Intelligence* do mercado.

Ela é um centralizador de ferramentas que habilitam as empresas a desenvolver soluções completas para resolver os problemas em *Business Intelligence*. Sua plataforma inclui componentes e relatórios para analisar os dados extraídos (PENTAHO CORPORATION, 2006).

Essa suíte contém ferramentas para executar todo o processo de construção de um *Data Warehouse*, e utilizar esse DW aplicando técnicas de *Business Intelligence*. Entre essas ferramentas podemos citar (PENTAHO CORPORATION, 2006):

- **Report Designer:** Ferramenta avançada para criação de relatórios. Possui grande flexibilidade e funcionalidades do que a capacidade de elaborar relatórios ad hoc que o BI Server Possui (PENTAHO CORPORATION, 2006).
- **Design Studio:** Ferramenta própria para editar na mão as visões de análise do BI Server. É usado geralmente para modificar relatórios feitos com Report Designer, acrescentando filtros em tempo de execução (PENTAHO CORPORATION, 2006).

- **Aggregation Design:** Cria metadados a partir de um *Data Warehouse* existente (PENTAHO CORPORATION, 2006).
- **Schema Workbench:** Uma ferramenta gráfica que ajuda na criação de esquemas ROLAP para análise (PENTAHO CORPORATION, 2006).
- **Pentaho Data Integration:** Ferramenta utilizada para extrair, transformar e carregar os dados de sistemas, planilhas ou arquivos de texto para o *Data Warehouse*. Também auxilia no mapeamento e criação do *Data Warehouse* (PENTAHO CORPORATION, 2006).
- **(Pentaho BI Server)** É um poderoso servidor on-line utilizado para acessar os *Data Warehouse* e gerar análises, relatórios e painéis de controle (PENTAHO CORPORATION, 2006).

Nesse trabalho utilizou-se duas dessas ferramentas para realizar o processo de *Business Intelligence*: O Pentaho Data Integration, foi utilizado para realizar os processos de ETL e o Pentaho BI Server foi utilizado para realizar as análises dos dados. Somente as duas ferramentas são suficientes para demonstrar o processo de Business Intelligence. As outras ferramentas apenas possuem recursos que auxiliam o processo, mas não imprescindíveis para o processo.

#### 2.4.1.1 Pentaho Data Integration

Essa ferramenta da Suíte Pentaho, também conhecida como Kettle, é um aplicativo poderoso capaz de extrair, transformar e carregar os dados dos sistemas em geral para um *Data Warehouse* (PENTAHO COMMUNITY, 2012).

A ferramenta é desenvolvida na linguagem de programação Java, e por isso pode ser utilizada em qualquer sistema operacional que utilize esse recurso e assim como todas as ferramentas da suíte Pentaho é *Open Source*. Trata-se de uma IDE projetada para realizar todo o processo de carga de dados em um *Data Warehouse* (PENTAHO COMMUNITY, 2012).

O Kettle é uma ferramenta gráfica com recursos de agarre e arraste muito intuitiva, e de fácil utilização. Com ela é possível se conectar a uma fonte de dados, tratar os dados extraídos e inseri-los em um *Data Warehouse* sem necessitar escrever nenhuma linha de código.

O Kettle consegue se conectar praticamente a qualquer fonte de dados, desde um simples arquivo de texto até um banco de dados robusto. Ele é capaz de se conectar a qualquer fonte de dados seja onde estiver (PENTAHO CORPORATION, 2012).

O Kettle é orientado a metadados, ou seja, é uma interface gráfica onde se especifica o que fazer sem escrever nenhuma linha de código. Para especificar o que se deseja basta adicionar etapas e conectá-las, para se formar uma seqüência de execução (GOMES, 2012).

Além de ferramenta de ETL ele pode ser utilizado em outras rotinas, como migração de dados entre servidores, exportar dados de bancos de dados para arquivos de texto, integração entre aplicações e manter a limpeza e qualidade dos dados num *Data Warehouse* (GOMES, 2012).

Seu funcionamento se dá através de etapas que são adicionadas e ligadas, formando um fluxo de execução. Cada uma dessas etapas são configuráveis, de acordo com a funcionalidade que representam.

As etapas são ligadas e formam um fluxo, porém isso não significa que uma etapa depende do final de execução de sua anterior para executar a próxima. Assim que cada etapa recebe dados da etapa anterior ela inicia sua execução, sendo executadas em paralelo (PENTAHO COMMUNITY, 2009).

#### **2.4.1.2 Pentaho BI Server**

O Pentaho BI Server é um servidor de gerenciamento para o processo de Inteligência Empresarial baseado em relatórios Web, fornecendo uma interface amigável com o usuário final e apresentando os resultados obtidos através de análises (PENTAHO COMMUNITY, 2012b).

Ele inclui elementos necessários para executar relatórios, análises, painéis gerenciais e mineração de dados. Ele possui níveis de acesso por usuários ou grupos de usuários, controlando o acesso às informações nele geradas e auditando tudo que nele é realizado (GUSMAN, 2012).

O Pentaho BI Server é responsável pela camada servidora do Pentaho. Ele oferece serviços que incluem logs, auditoria, serviços web, motores de regras de negócio, mecanismos para integração com relatórios, análises e painéis gerenciais (AMBIENTE LIVRE, 2012).

#### **2.4.2 Sistema Acadêmico ADX**

O AdX é um sistema desenvolvido para dar suporte a todo o processo acadêmico de uma instituição de ensino. Ele possibilita a automação de todo o processo acadêmico, e controle de recebimentos (FLUX SOFTWARES, 2012).

O sistema AdX gerencia informações para secretaria, biblioteca, tesouraria, professores, alunos, coordenadores e diretoria, integrando todas as áreas e fornecendo vantagens como acesso Web e automação de diversas tarefas acadêmicas (FLUX SOFTWARES, 2012).

Por se tratar de um sistema com acesso Web, sua versatilidade permite que os usuários acessem o mesmo onde quer que estejam desde que possuam uma conexão com a Internet.

Para garantir a segurança dos dados, o sistema de acesso é realizada em duas camadas, sendo a primeira garantida pelo próprio sistema e segunda feita pelo banco de dados (FLUX SOFTWARES, 2012).

O módulo de tesouraria permite ao usuário configurar as opções de recebimento, permitindo o controle dos boletos, mas flexibilizando as opções de geração e disponibilização dos mesmos.

O sistema garante a eficiência da operação através de rotinas integradas, como o lançamento de notas exibidas aos alunos e a geração de boletos disponibilizados na interface do aluno (FLUX SOFTWARES, 2012).

O sistema é desenvolvido com a linguagem de programação PHP e utiliza bancos de dados MySQL. Cada filial possui uma base de dados separada altamente normalizada, tornando o sistema totalmente confiável.

Segundo ELMARI e NAVATHE (2005) o processo de normalização garante que o esquema relacional utilizado em um banco de dados satisfaça três normas: não haja repetição de dados, pois cada atributo é atômico e mono valorado, que não haja dependências funcionais não-triviais de atributos não chaves a não ser deles com a chave primária, e o último diz que além de respeitar a primeira e segunda norma também não deve haver dependência transitiva entre os atributos. Isso garante que o banco de dados funcionará corretamente, mantendo um armazenamento confiável dos dados e fazendo com que as transações diárias ocorram da melhor maneira possível.

### 3 METODOLOGIA

Para realizar esse trabalho utilizou-se os bancos de dados do sistema acadêmico AdX. Cada filial da instituição de ensino possui um banco de dados separado para cada unidade de negócio. Por exemplo, existe um banco de dados para a graduação de Caratinga, outro para o Colégio de Caratinga e outro para a Graduação de Cataguases. Apesar dos bancos de dados serem altamente normalizados, garantindo a eficácia da operação não existe relatórios consolidados para se comparar os dados das unidades de forma conjunta.

O Trabalho consistiu em analisar esses bancos de dados, para extrair, transformar e unificar esses dados em um único local, um *Data Warehouse*, para que o acesso a eles fosse realizado mais facilmente. Com esses novos dados, empregou-se uma ferramenta OLAP para análise da informação, demonstrando a flexibilidade e a rapidez na elaboração dos comparativos.

Na seção 3.1 será mostrado a análise dos bancos de dados do AdX, descrevendo as tabelas que foram utilizadas para a análise financeira. Na seção 3.2 será mostrado o modelo dimensional criado, explicando suas dimensões. Na seção 3.3 será mostrado como foi realizado o processo de povoamento do *Data Warehouse*, detalhando como foi criado o processo automatizado para povoar cada dimensão e tabela fato. Na seção 3.4 será mostrado como os dados armazenados no *Data Warehouse* foram modelados para gerar análises, e como os relatórios, cubos OLAP e painéis de controle são criados.

#### 3.1 ANÁLISE DO SISTEMA ADX

Segundo OLIVEIRA (2006), para se desenvolver um projeto de *Data Warehouse* de qualidade é necessário focar na área que se deseja explorar e extrair dados de qualidade, pois essa etapa influenciará todas as demais.

O sistema AdX possui instâncias para cada filial e cada unidade de negócio da instituição de ensino. Cada instância possui um banco de dados distinto.

Todos os bancos são padronizados e altamente normalizados para garantir a eficácia da operação transacional diária.

Para desenvolvimento desse trabalho foram extraídas informações do sistema AdX referentes à vida financeira dos alunos, mais especificamente ao faturamento – os boletos emitidos. Para isso, estudou-se as tabelas que se relacionavam para formar as informações relevantes à essa informação financeira como dados pessoais do aluno, período letivo a que pertence o boleto e os dados das matrículas dos alunos nos períodos letivos.

A Figura 7 representa as tabelas e suas relações, de um dos bancos de dados do sistema AdX utilizadas para extrair as informações financeiras.

As tabelas mapeadas foram:

- tes\_boleto;
- tes\_tipo\_pgto;
- acad\_aluno;
- acad\_matricula;
- acad\_item\_matricula;
- ped\_curso;
- periodo\_letivo;

Essas seis tabelas contêm os dados necessários para uma análise financeira, sendo eles:

- Período letivo;
- Curso do aluno;
- Turma do aluno no período letivo;
- Período cursado pelo aluno no período letivo;
- Registro acadêmico do aluno;
- Dados pessoais do aluno;
- Valor emitido do boleto;
- Data de emissão do boleto;
- Data vencimento do boleto;
- Parcela que se refere o boleto no período letivo;
- Valor pago pelo aluno;
- Tipo de pagamento realizado;

- Data de pagamento do boleto;

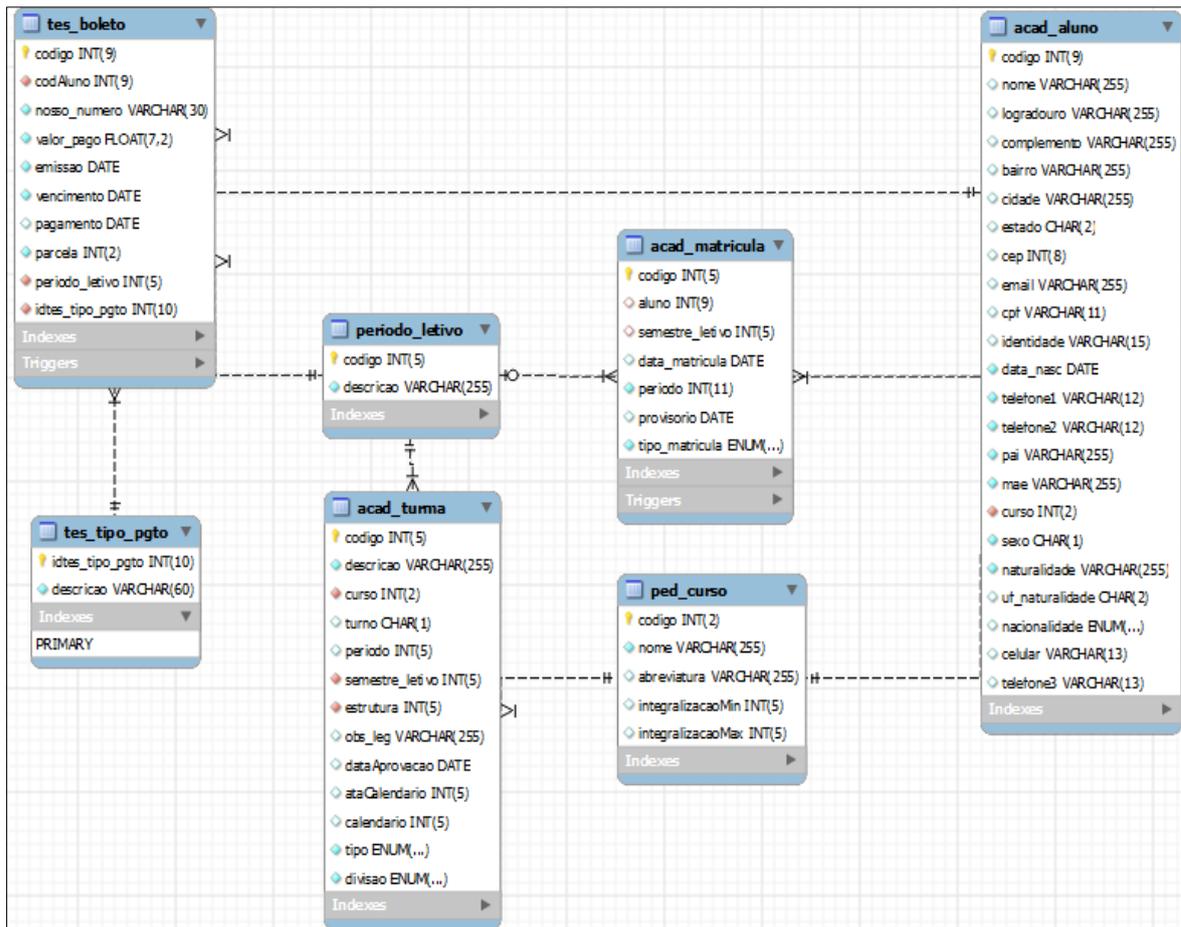


Figura 7: Modelo Relacional – Relacionamento entre Tabelas geradoras de informações financeiras do AdX (AdX).

Além de identificar essas informações, constatou-se que duas informações estavam implícitas neles, e para o projeto elas deveriam ser identificadas e mapeadas: a filial a que as informações pertencem, e a unidade de negócio. Como os dados de cada unidade de negócio (graduação, colégio, pós graduação, mestrado) e cada filial se encontram em bancos de dados separados e serão colocados em apenas um local, deve existir um identificador que permita classificar os dados quanto a esses quesitos.

Essas informações não estão armazenadas em nenhum local, porém no momento da extração elas foram inseridas durante o processo de ETL.

### 3.2 CRIANDO O DATA WAREHOUSE

Foi utilizada a maior granularidade permitida pelo banco de dados do sistema AdX, ou seja, será possível analisar boleto a boleto, todos os pagamentos efetuados por um aluno. Esse aluno por sua vez estará associado a uma turma, dentro de um período, curso, unidade de negócio e filial.

Segundo CRAMER (2006), o desempenho das consultas no *Data Warehouse* está diretamente ligada ao nível de granularidade utilizada. Ou seja, quanto maior a granularidade, melhor o desempenho. Porém, uma baixa granularidade permite uma melhor análise no nível mais baixo da informação.

O termo *Data Warehouse* é utilizado para definir o conjunto de informações coletadas de uma ou diversas fontes, tratadas e armazenadas em um único local, seguindo determinado esquema. Dessa forma, os dados armazenados permitem acesso de forma rápida a dados históricos (SILBERSCHATZ et al., 2006).

Para que os dados sejam armazenados e resgatados de forma eficaz, eles foram armazenados em um *Data Warehouse*. No modelo dimensional desse projeto, utilizou-se o esquema estrela, pois segundo CRAMER (2006) esse esquema propicia o desempenho, tornando o modelo mais simples.

Criou-se uma tabela fato para armazenar os dados principais, como período letivo, datas e valores dos boletos. Essa tabela é a principal e faz ligação com as dimensões.

As dimensões criadas foram a dimensão data, dimensão aluno e dimensão boleto, conforme esquema da Figura 8. Esse modelo representa o *Data Mart* criado para trabalhar com os dados financeiros, mais especificamente, com os recebimentos da empresa.

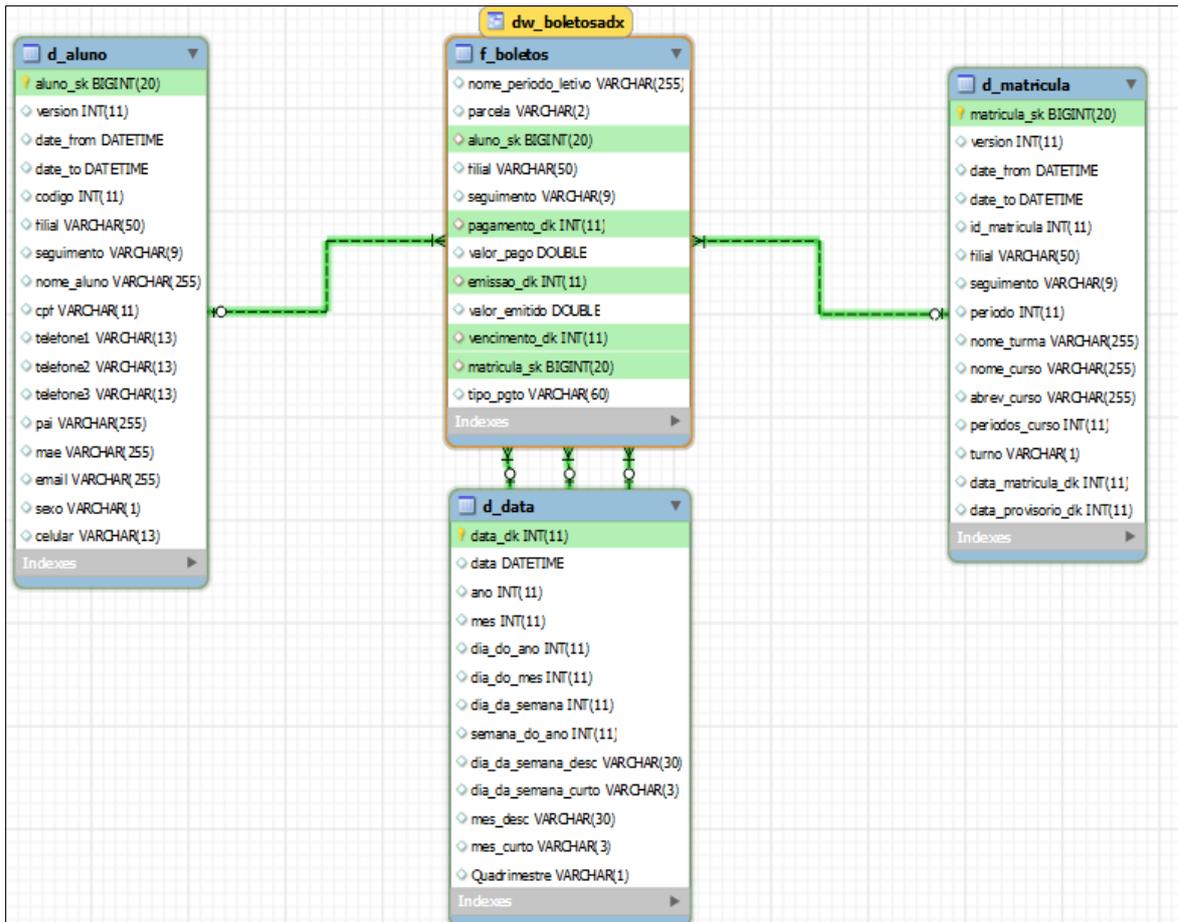


Figura 8: Modelo Dimensional - Esquema Estrela criado para o Data Warehouse

A tabela fato `f_boletos` é a tabela fato e conseqüentemente contém as medidas a serem utilizadas no projeto, como valor emitido e valor pago. Além desses dados, ela contém as chaves delegadas para as tabelas dimensões.

A tabela fato é a mais importante do projeto. Ela contém as informações quantitativas, e outras que podem ser extraídas a partir das já existentes, como um contador para o número de boletos pagos, ou o valor pago fora do mês em determinada parcela.

A tabela `f_boletos` faz ligação com a dimensão `data` através das chaves delegadas `emissao_dk`, `vencimento_dk` e `pagamento_dk`. A ligação com a dimensão `matricula` foi realizada através da chave delegada `matricula_sk`. Já a ligação com a dimensão `aluno` é realizada através da chave delegada `aluno_sk`.

A tabela dimensão `d_data` é uma dimensão indispensável a qualquer *Data Warehouse*. Ela contém informações relacionadas com a data armazenada. Essa

dimensão é muito útil, pois contém além de datas, dados relacionadas a elas, como o dia da semana a que se refere, a qual trimestre se refere, e várias outras, que são representados em campos separados, facilitando sua utilização, e diminuindo a complexidade na sua utilização (MACHADO, 2007).

Nesse projeto ela é constituída de cerca de quarenta mil registros, compostos por datas desde o ano de mil novecentos e cinquenta, a dois mil e cinquenta e nove. Isso se deu para que todas as datas presentes no sistema AdX possuam correspondente nessa tabela. Esse intervalo permite representar todos os períodos de emissão, vencimento e pagamento dos boletos, garantindo que elas estarão representadas nessa tabela.

A tabela dimensão d\_aluno contém os dados pessoais dos alunos. Ela guarda os registros de todos os alunos extraídos do sistema AdX. Ela possuirá um identificador único para cada aluno da rede, independente se o número de registro em mais de uma unidade for conflitante.

A tabela dimensão d\_matricula é a junção de algumas informações, como período letivo, nome do curso, turma e período letivo referente àquela matricula. O resultado obtido dessa junção, armazenado dessa forma faz com que o resgate das informações seja mais ágil.

### **3.3 POVOANDO O DATA WAREHOUSE**

Para extrair, transformar e carregar os dados utilizou-se a ferramenta Kettle. Ela contém funcionalidades que facilitam o processo. Para tanto, criou-se um trabalho que foi executado diariamente, atualizando os dados de pagamento dos alunos em tempo real.

Os dados foram extraídos diretamente do sistema AdX. Foram adicionadas conexões para comunicação direta com os bancos de dados de cada unidade conforme representado na Figura 9. Foram utilizados dezenove bancos de dados do sistema AdX, de onde as informações foram extraídas, tratadas e armazenadas em um único local para análise de dados.

Para povoar o *Data Warehouse* foram necessárias três transformações, repetidas em cada um dos dezanove banco de dados do AdX. Essas transformações são responsáveis por povoar as dimensões aluno, matrícula e boletos.

A dimensão data foi carregada apenas uma vez, pois seus dados não sofrem alteração. A própria ferramenta Kettle contém uma transformação pré-definida para carregar a dimensão data. Uma vez executada, ela não necessita de atualizações.

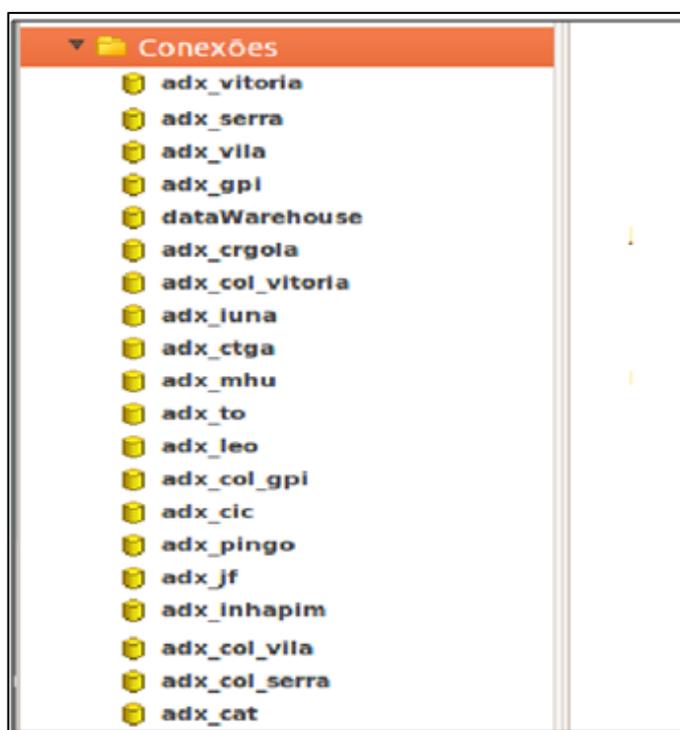


Figura 9: (Pentaho Kettle) Conexões utilizadas no processo de ETL

### 3.3.1 Dimensão Alunos

A transformação para gerar a dimensão alunos é a mais simples entre as três transformações que foram construídas. Ela consiste em ler os dados pessoais dos alunos, e armazená-los no *Data Warehouse*.

A chave primária da dimensão alunos é gerada levando-se em conta três dados: o código do aluno, a unidade de negócio e a filial do aluno. Ao salvar os dados dos alunos essas informações foram armazenadas junto com eles, pois serão utilizadas para povoar a tabela fato e realizar o mapeamento dos alunos.

Esse cuidado foi tomado porque um código do aluno pode ser o mesmo para alunos de filiais diferentes. Para realizar um mapeamento correto esse cuidado deve ser tomado.

As etapas da transformação, conforme mostrado na Figura 10 são:

- **Leitura dos dados:** Essa etapa é responsável por realizar a conexão com o banco de dados do AdX. Nela é configurada a conexão a ser utilizada, e uma consulta para leitura da tabela acad\_alunos. Nessa tabela estão os dados pessoais dos alunos. Na consulta foi inserida a filial e a unidade de negócio a que se refere à conexão.
- **Gravação dos dados:** Após a extração e mapeamento dos dados, eles são armazenados no *Data Warehouse*. Para isso é configurado a conexão com o *Data Warehouse* e a tabela que será povoada.

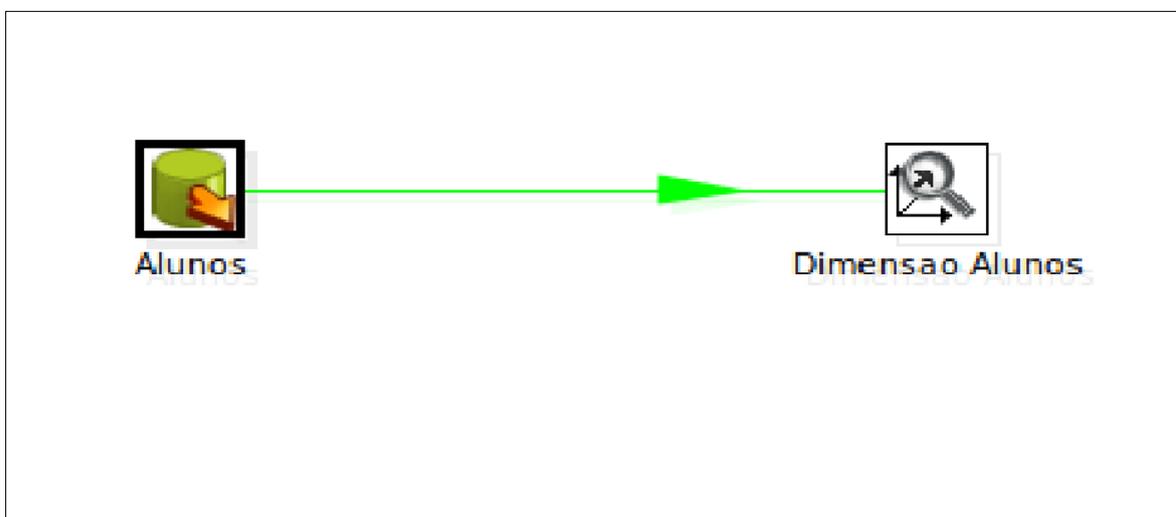


Figura 10: (Pentaho Kettle) Geração da Dimensão Alunos

### 3.3.2 Dimensão Matricula

A dimensão matrícula requereu a leitura de várias tabelas, que unidas forneceram a informação de matrícula do aluno. Nela foram utilizadas quatro tabelas e a junção entre elas forneceu os dados da matrícula.

As etapas dessa transformação são listadas na Figura 11. Abaixo segue a descrição para cada uma das etapas.

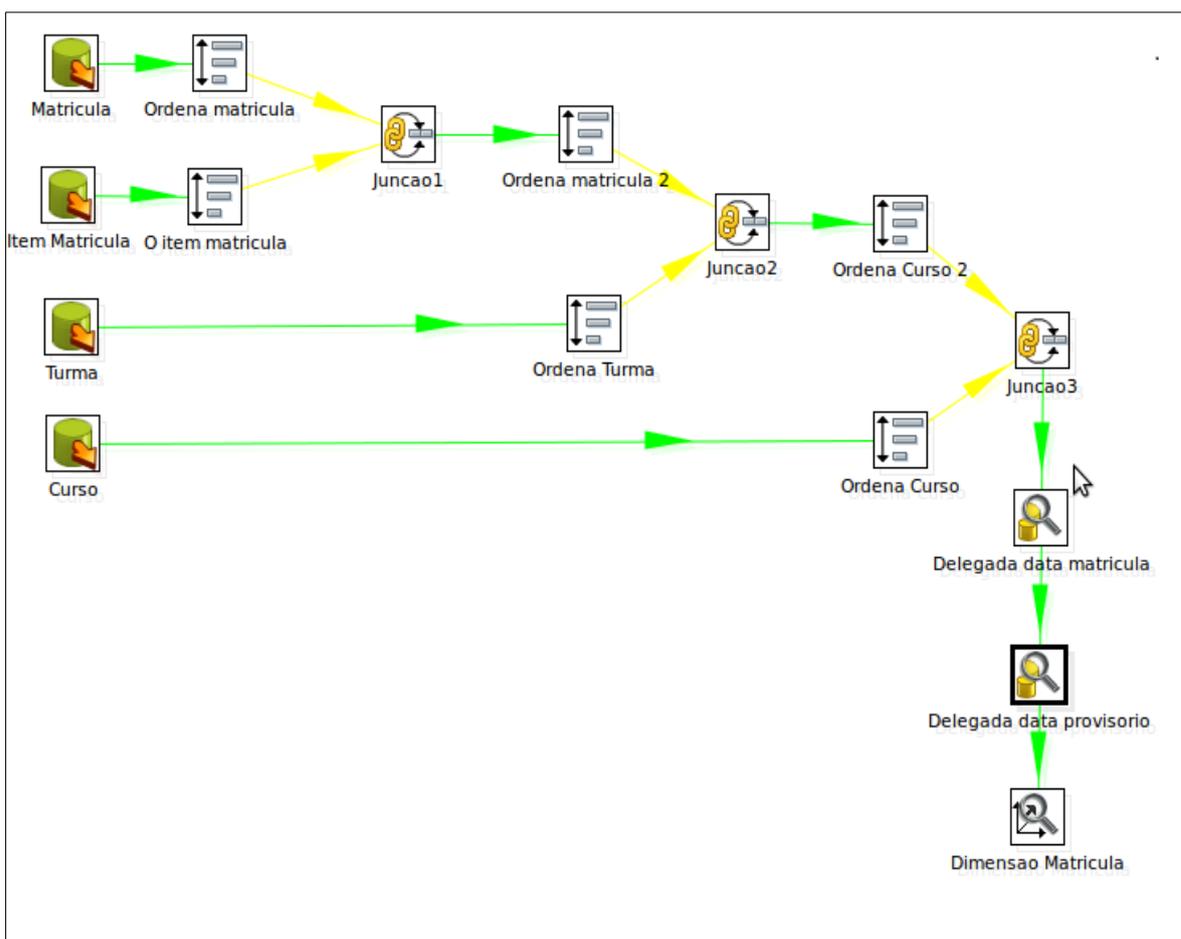


Figura 11: (Pentaho Kettle) Geração da Dimensão Matrícula

- **Matrícula, item matrícula, turma e curso:** Etapas responsáveis por ler as tabelas com esses respectivos nomes. Em cada uma dessas etapas é configurada uma conexão com o banco de dados do AdX e uma consulta simples responsável pela leitura da respectiva tabela. Os dados extraídos dessas tabelas serão unidos nas próximas etapas.

Na etapa matrícula é adicionada a unidade a que os dados se referem e a unidade de negócio, para ser utilizados na confecção da chave delegada na tabela fato.

- **Ordena matrícula e ordena item matrícula:** Etapas indispensáveis para realizar junção entre tabelas. A ferramenta realiza junção entre tabelas desde que o campo utilizado para a junção esteja ordenado.
- **Junção um:** Contém os parâmetros necessários para unir as tabelas. Recebe o campo da primeira e da segunda tabela que serão utilizados para uni-las. Nessa etapa é recebido o código da turma a que a matrícula pertence.
- **Ordena matrícula dois, e ordena turma:** Ordena o resultado da última junção e os dados da tabela turma para serem unidos.
- **Junção dois:** Une os dados recebidos das etapas anteriores, utilizando critérios nele configurados. Do resultado dessa junção sairá o nome da turma referente à matrícula.
- **Ordena curso, e ordena curso dois:** Etapa responsável por ordenar os dados recebidos da última junção, e os dados extraídos da tabela de cursos para realizar a próxima junção.
- **Junção três:** Une os dados recebidos da ordenação anterior. Etapa responsável por adicionar o nome do curso à matrícula. É a última junção realizada na transformação.

A partir daí todos os dados necessários estão unidos. As próximas etapas consistem em criar as chaves delegadas para as datas. A substituição das datas por chaves delegadas referentes à dimensão data enriquecerá a informação com dados complementares às datas.

- **Chave delegada data matrícula, e data provisório:** Etapas responsáveis por criar as chaves delegadas. Ela procura a data de matrícula e a data de matrícula provisória dos dados recebidos na tabela dimensão data e retorna a chave primária da tabela dimensão data para ser armazenada na dimensão matrícula.
- **Dimensão matricula:** Essa etapa realiza a conexão com o *Data Warehouse* e a tabela que será utilizada para armazenar as informações da dimensão.

Também é responsável por criar a chave primária da tabela e filtrar quais campos serão inseridos nela. Essa etapa recebe os dados resultantes das uniões e também os criados para chave delegada.

### 3.3.3 Tabela Fato

Para se criar a tabela fato, foi necessário ler quatro tabelas. Além de unir a tabela de períodos letivos a esse processo, também foi necessário incluir a tabela de matrículas, para se criar a chave delegada matrícula.

Como pode ser observado na Figura 12, é a transformação mais complexa, pois além de extrair os dados do AdX e armazená-los no *Data Warehouse*, ela cria chaves delegadas para as dimensões carregadas anteriormente.

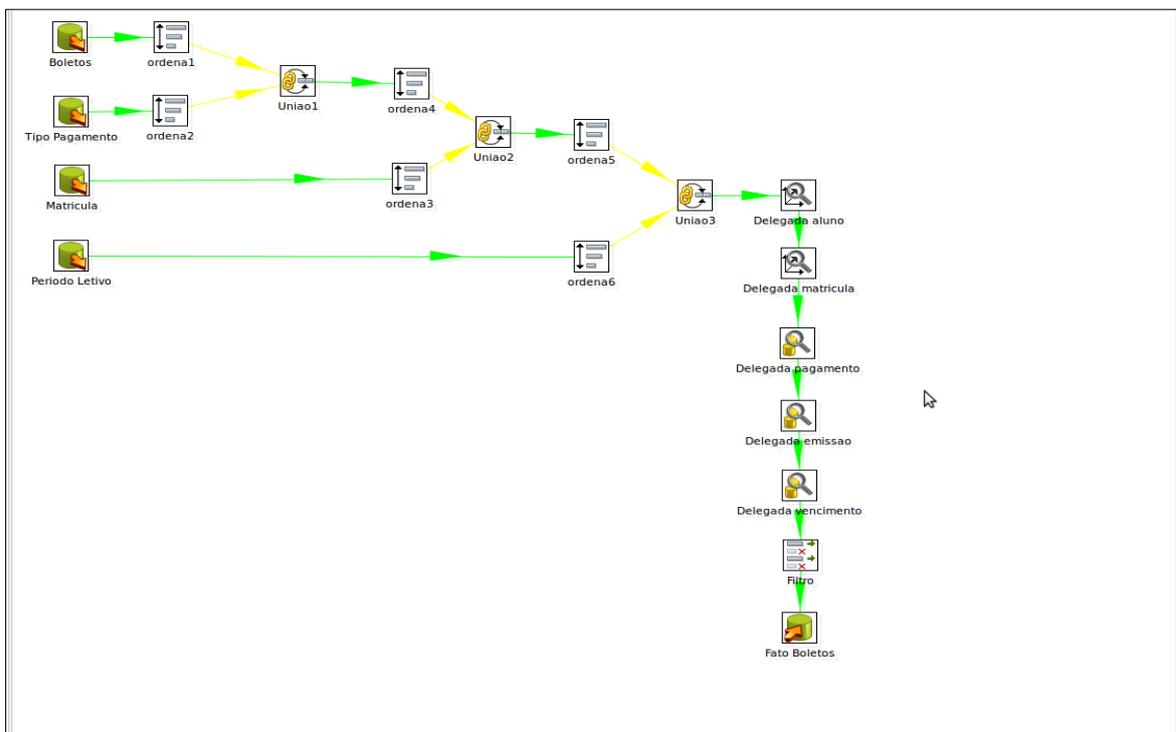


Figura 12: (Pentaho Kettle) Geração da Tabela Fato

Nessa transformações são encontradas as seguintes etapas:

- **Boletos, tipo pagamento, matrícula e período letivo:** Etapas que fazem conexão com o banco de dados do AdX e lêem os dados dessas tabelas. Esses dados serão unidos nas próximas etapas.  
A extração mais importante é a de boletos, porém a união com as outras tabelas se faz necessária para acrescentar dados importantes para o processo. Na etapa boletos, além dos dados dos boletos, foi acrescentado a filial a que pertence o banco de dados do AdX e a unidade de negócio da mesma.
- **Ordena um e ordena dois:** Etapas indispensáveis para o processo. O Kettle realiza a junção entre duas tabelas somente se elas estiverem ordenadas pelo campo a ser unido. Essa etapa ordena os dados das etapas a que está ligada pelo campo tipo de pagamento.
- **União um:** Une os dados recebidos das etapas anteriores através do campo tipo de pagamento. Essa união acrescenta aos dados dos boletos o tipo de pagamento, caso estejam pagos.
- **Ordena três, e ordena quatro:** Novamente é realizada a ordenação entre o resultado da última junção e os dados da etapa matrícula. Os dados são ordenados pelo campo de código de aluno e código do período letivo que serão utilizadas na próxima junção.
- **União dois:** Após ordenados, os dados são unidos, e dessa forma os boletos ficam relacionados a uma matrícula dos alunos.
- **Ordena cinco e ordena seis:** Outra ordenação, dessa vez pelo campo código do período letivo, para que a informação do período letivo seja adicionada aos boletos.
- **União três:** Essa união acrescenta o nome do período letivo aos boletos. Com essa junção, todos os dados necessários estão gerados. As etapas seguintes são responsáveis por fazer a ligação entre a tabela fato e as dimensões carregadas anteriormente.
- **Chave delegada aluno:** Para gerar a chave delegada aluno foi utilizado três campos – filial, unidade de negócio e código do aluno – que juntos formam uma informação única.

Essa etapa se conecta a dimensão aluno, compara os três campos com os homônimos da dimensão e retorna a chave primária, que será armazenada como chave delegada.

- **Chave delegada matrícula:** Da mesma forma que foi gerada a chave delegada aluno foi gerada a chave delegada matrícula. Essa etapa realiza a conexão com a dimensão matrícula.

Os três campos: filial, unidade de negócio e código e matrícula, foram comparados aos seus homônimos na dimensão matrícula. A etapa retorna então a chave primária da dimensão que é utilizada como chave delegada na tabela fato.

- **Delegada pagamento, delegada emissão e delegada vencimento:** Para enriquecer os dados, as datas de pagamento, emissão e vencimento foram substituídas pela chave primária da dimensão data.

Dessa forma foi inserida na tabela fato as chaves delegadas para emissão, vencimento e pagamento dos boletos. Após criar as chaves delegadas os dados passaram por um filtro para serem salvos na tabela fato.

- **Filtro:** Um filtro foi aplicado para selecionar os campos a ser salvos na tabela fato. Essa etapa recebe todos os campos gerados pelas etapas anteriores. Nesse momento são selecionados apenas campos que irão agregar valor ao modelo. Os campos de data, código de aluno e código de matrícula não foram selecionados, pois eles serão substituídos por chaves delegadas.

- **Tabela fato boletos:** Após filtrar os dados, os mesmos foram salvos na tabela fato boletos. Essa etapa é responsável por se conectar ao *Data Warehouse* e configurar a tabela a ser utilizada para armazenar os dados aqui extraídos e tratados.

### 3.3.4 Criando Jobs (Trabalhos) e Encapsulando as Transformações

As transformações descritas na seção anterior foram repetidas para cada um dos dezenove bancos de dados do sistema AdX utilizados. Cada transformação foi salva em um arquivo com formato próprio para ser executado pelo Kettle. Foram gerados três arquivos por instância do AdX, totalizando cinquenta e sete arquivos para esse trabalho. O Kettle possui outro tipo de arquivo chamado *job*, onde pode-se criar um fluxo de execução incluindo várias transformações, ou até mesmo outros *jobs*.

Assim, foi criado um *job* para cada unidade, cada um contendo um fluxo de execução com as três transformações criadas para carregar a dimensão aluno, dimensão matrícula e tabela fato, conforme pode ser visto na Figura 13.

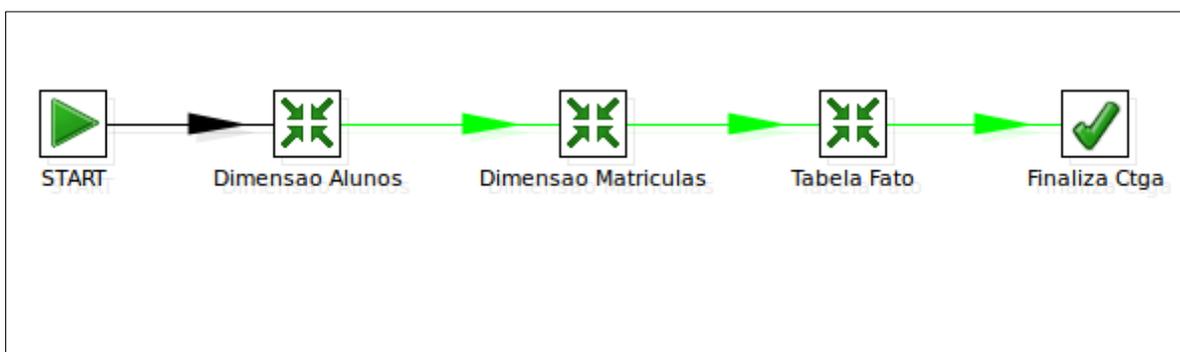


Figura 13: (Pentaho Kettle) Encapsulamento de Transformações em um único Job

Cada *job* é composto pelas três transformações criadas, aplicadas a cada unidade. A transformação responsável pela tabela fato deve ser a última a ser realizada, pois depende das outras dimensões carregadas para ser gerada.

A etapa inicial e final foram inseridas apenas para complementar o processo. Elas são obrigatórias para delimitar o trabalho e sem elas o trabalho não é executado.

### 3.3.5 Criando o Arquivo Principal

Para que os dados de todas os bancos de dados utilizados fossem extraídos de uma vez, foi necessário encapsular todos os *jobs* criados em um único, responsável por executar toda a sequência.

Na Figura 14 podemos ver todos os *jobs* já encapsulados em um único, na sequência em que foram executados. Além dos *jobs* responsáveis pela extração de cada unidade, foram adicionadas a esse trabalho principal outras etapas importantes para o processo.

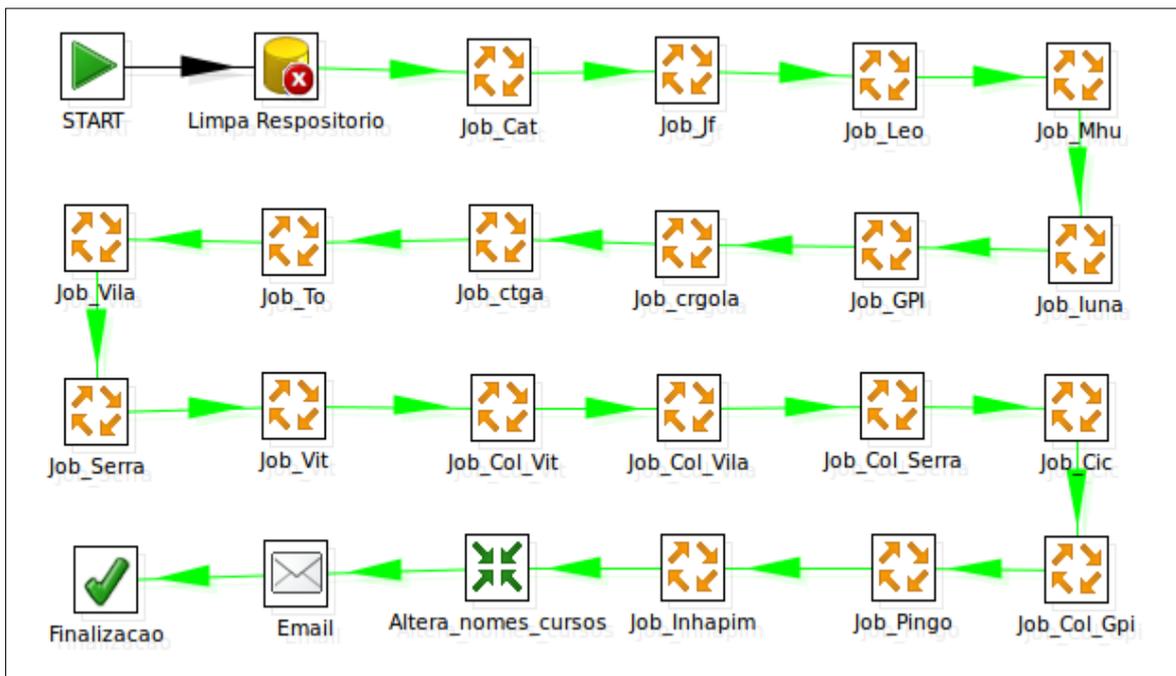


Figura 14: (Pentaho Kettle) Encapsulamento de Jobs em um único Job.

A etapa “limpa repositório” é a responsável por limpar os dados da última carga realizada no *Data Warehouse*. A cada execução do processo de carga, essa etapa é a primeira a ser realizada. Como o próprio símbolo indica, ela exclui os dados do *Data Warehouse* para que novos dados sejam adicionados.

Após ser realizada a carga dos dados de todas as instâncias do sistema AdX, é executada uma etapa para alterar nomes de alguns cursos. Isso foi necessário para conseguir uma homogeneidade na nomenclatura dos cursos, visto que alguns nomes estão com divergentes de uma filial para outra. A etapa final envia um email para o responsável pela rotina, informando o final do processo.

Segundo MACHADO (2007), o tratamento de conflitos, correção de nomenclaturas não normalizadas deve ser feita nessa etapa, para que os dados já estejam corretos ao se realizar as consultas nos mesmos.

Esse *job* pode ser agendado para executar com determinada frequência tornando o processo automático.

### **3.4 MODELANDO OS DADOS DO *DATA WAREHOUSE***

Após povoar o *Data Warehouse*, a próxima etapa foi modelar os dados para serem utilizados em relatório configurável e nos cubos OLAP. Dessa forma podem-se unir as dimensões à tabela fato para analisar os dados sob diferentes óticas.

A partir dessa etapa foi utilizado o Pentaho BI Server para modelar os modelos de dados e exibir os resultados obtidos. Ele permite a conexão com todos os bancos de dados existentes no mercado (PENTAHO CORPORATION, 2006).

Para criar um novo modelo existe uma interface no Pentaho BI Server para adicionar fontes de dados. A Figura 15 mostra a primeira etapa dessa interface, onde deve ser selecionada uma conexão e a consulta que irá gerar os metadados. Foi criada uma conexão com o *Data Warehouse* para ter acesso às tabelas. Deve ser selecionado o tipo de conexão, banco de dados a ser utilizado, endereço do servidor de bancos de dados, o usuário e senha do banco de dados, para se conectar à fonte de dados.

Foi selecionado o tipo de conexão para banco de dados MySQL e configurado o servidor local onde está hospedado o banco de dados. Foi selecionado o nome de bancos de dados *dw\_boletosadx* que é o banco onde se encontram os dados consolidados.

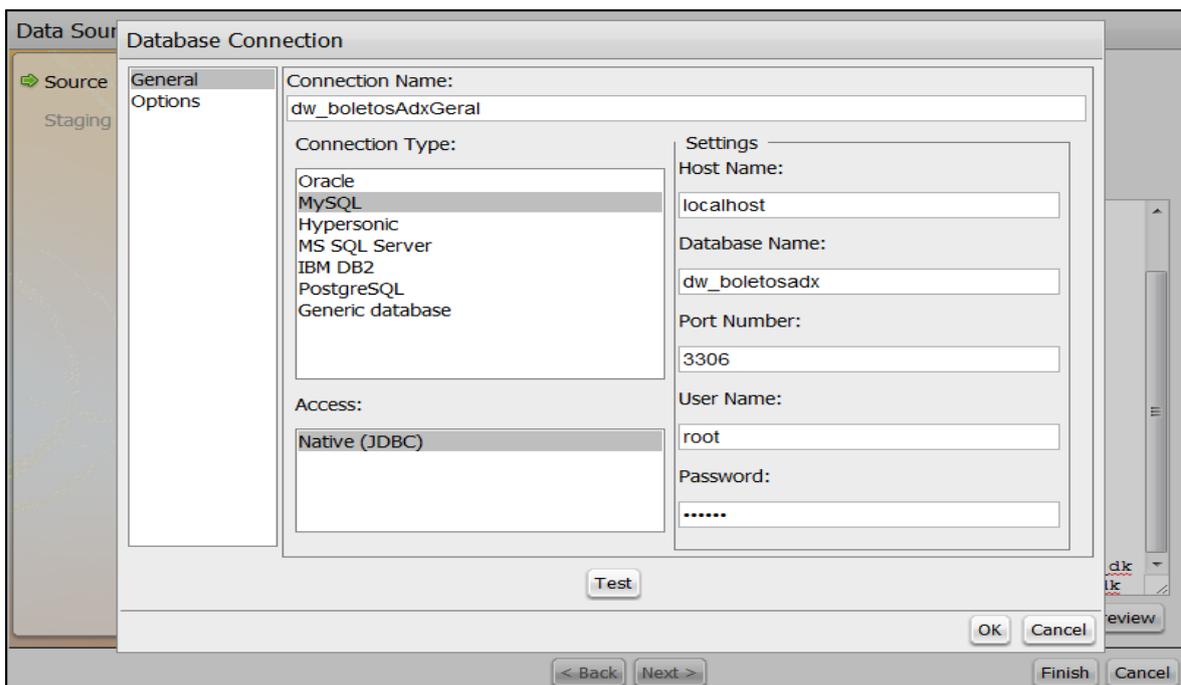


Figura 15: (Pentaho BI Server) Modelo de Dados - Configurando Conexão

A próxima etapa consistiu em manipular os campos resultantes da conexão. Foi elaborada uma consulta SQL unindo a tabelas fato às dimensões, gerando um modelo de dados que resultou nas seguintes colunas: aluno, curso, data de emissão, filial, número de boletos pagos, numero total de boletos, data de pagamento, valor pago, parcela a que se refere o boleto, período de matricula do aluno, período letivo, unidade de negócio do aluno, nome da turma, valor emitido e vencimento do boleto, conforme se pode ver na Figura 16. O BI Server também mapeou os campos numéricos, sugerindo que fossem utilizados como campos comparativos, ou simplesmente medidas.

A interface também permite que os campos sejam formatados e sejam adicionadas funções para contabilizar os totais das medidas. Após formatar os campos conforme necessário, o modelo foi salvo.

O nome dos campos pode ser alterado, deixando o modelo mais fácil de ler. A ordem dos campos também pode ser alterada, bem como excluídos. Após modelar os dados conforme necessário, o modelo ficou como apresentado na Figura 16.

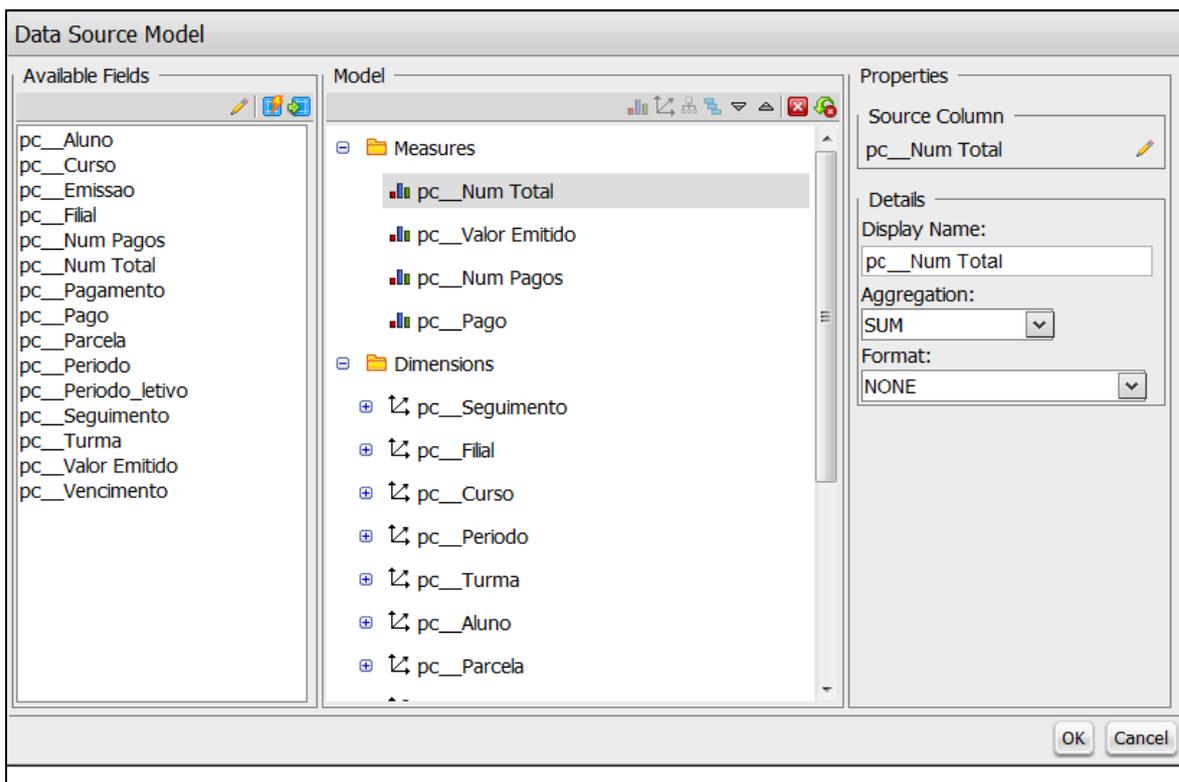


Figura 16: (Pentaho BI Server) Modelo de Dados - Fase Final

O modelo criado possui quatro campos que podem ser medidos: número total de boletos, valor emitido, número de boletos pagos e valor pago. Esses campos fornecerão totais acerca da informação analisada.

Todos os outros campos podem ser ordenados e parametrizados de forma necessária a obter a melhor informação. A utilização desse modelo será tratada na próxima seção.

### 3.4.1 Utilizando a Informação Gerada

Com os dados salvos no *Data Warehouse* e com um modelo de dados criado, a próxima etapa foi extrair relatórios e análises de dados obtidos através do modelo.



- **Seleção dos campos:** Essa etapa é uma das principais da criação do relatório. Conforme Figura 18, pode-se ver que a interface possui recursos para agrupar os dados em diferentes níveis. Os campos do modelo são exibidos na seção de itens disponíveis. Para adicioná-los ao relatório basta arrastá-los e soltá-los na janela de detalhes. Os campos serão exibidos na ordem que estiverem nessa janela. Na seção agrupamento deve ser adicionado o campo que será utilizado para agrupar os dados, gerando nível no corpo do relatório. Na seção filtro devem ser adicionados os campos aos quais serão aplicados filtros para analisar determinado período ou produto.

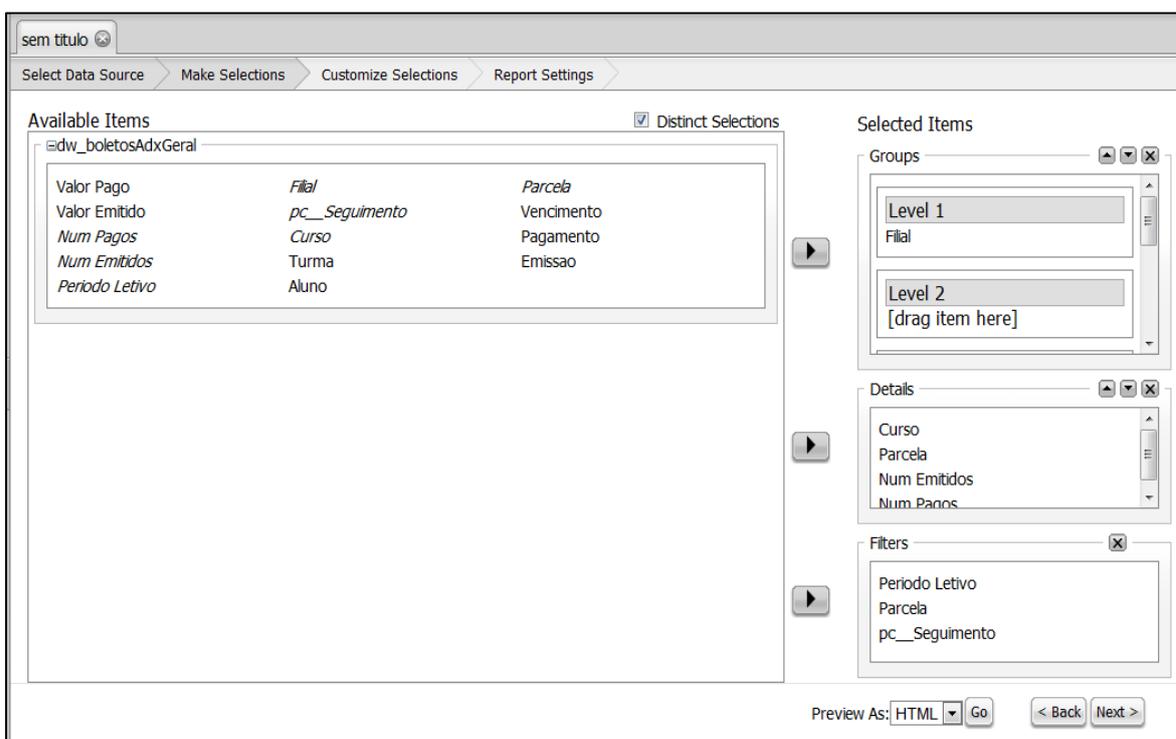


Figura 18: (Pentaho BI Server) Relatório - Selecionando Campos.

- **Aplicação de Filtros e Formatação de campos:** Após selecionar os campos a ser detalhados, agrupados e filtrados, nessa etapa o usuário pode aplicar filtros e formatar os campos do relatório. A Figura 19 mostra a interface, frisando a seção de filtros. Clicando em um campo na seção filtros é exibida a seção para adicionar os filtros. Clicando em algum campo na seção detalhes ou agrupamento, é

exibida a seção para formatar os campos e outra para associar funções de soma ou contador aos mesmos.

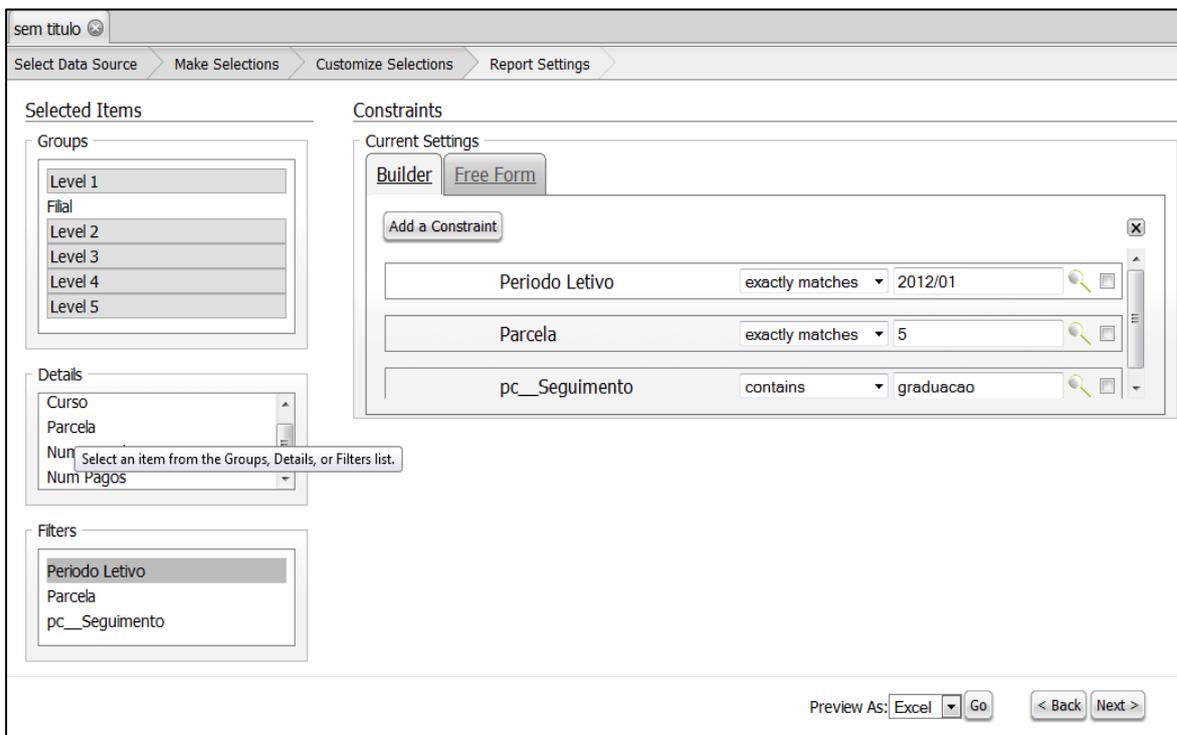


Figura 19: (Pentaho BI Server) Relatório - Aplicando Filtros

- **Configurações do Relatório:** Nessa etapa são realizadas configurações gerais do relatório, como orientação, cabeçalho e rodapé da página, tamanho do papel, títulos do relatório e da página e o tipo de relatório a ser gerado. Entre os tipos de relatório pode ser definido o tipo HTML para exibição no próprio servidor, ou ainda os formatos PDF, Excel, e CSV. A Figura 20 ilustra essa etapa.

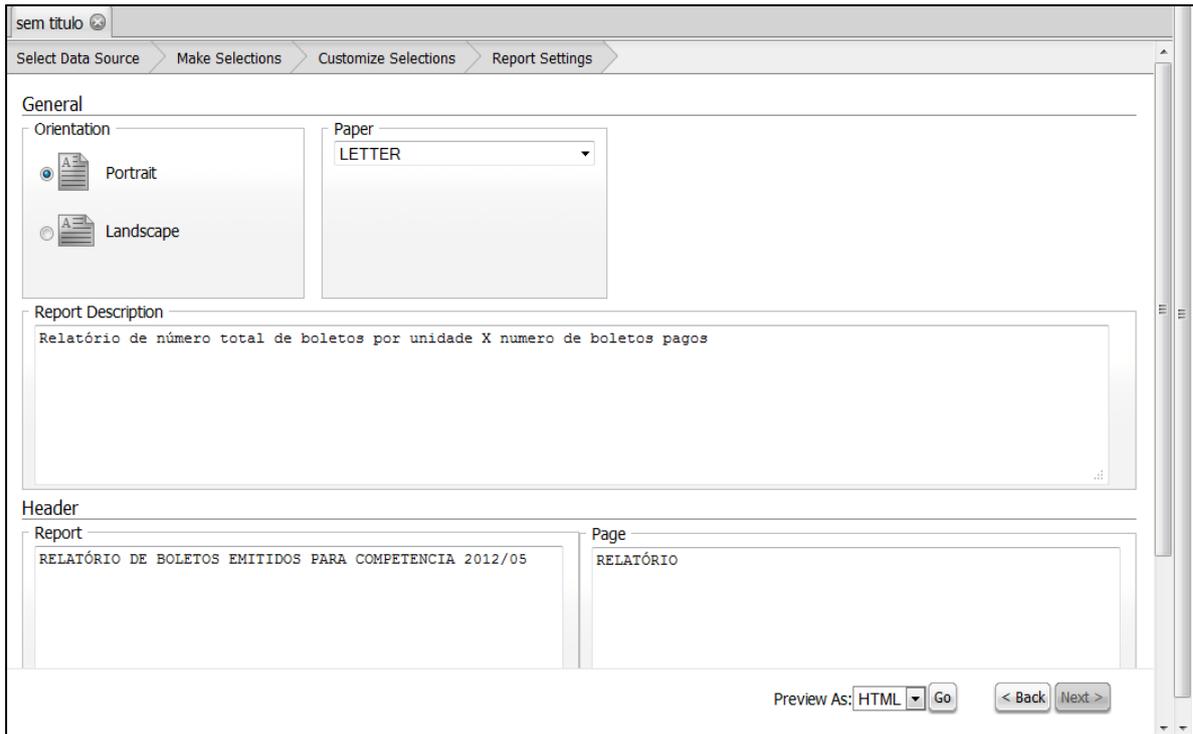


Figura 20: (Pentaho BI Server) Relatório - Configuração de Página

- **Exibindo o Relatório:** Após definir todos esses parâmetros o relatório estará pronto para ser gerado. Na Figura 21 pode-se ver o relatório exibido em HTML no próprio servidor. Nele é possível ver os agrupamentos e subtotais aplicados a eles.

RELATÓRIO DE BOLETOS EMITIDOS PARA COMPETENCIA 2012/05			
Filial: Carangola			
Curso	Parcela	Num Emitidos	Num Pagos
Direito	5	282	205
<b>Total Carangola</b>		<b>282</b>	<b>205</b>
Filial: Caratinga			
Curso	Parcela	Num Emitidos	Num Pagos
Ciência da Computação	5	69	61
Ciências Contábeis	5	183	164
Direito	5	576	498
Engenharia Civil	5	397	330
Engenharia Elétrica	5	107	91
Serviço Social	5	76	71
<b>Total Caratinga</b>		<b>1408</b>	<b>1215</b>
Filial: Cataguases			
Curso	Parcela	Num Emitidos	Num Pagos
Enfermagem	5	41	35
Sistemas de Informação	5	88	82
<b>Total Cataguases</b>		<b>129</b>	<b>117</b>

Figura 21: (Pentaho BI Server) Relatório de boletos por Filial

Essas etapas apresentam a criar um novo relatório utilizando um modelo de dados adicionado ao servidor. Para deixar esse relatório permanente, ele deve ser salvo em uma pasta do servidor, para futuras consultas.

### 3.4.1.2 Cubo OLAP

Os cubos são a forma mais dinâmica de analisar os dados armazenados no *Data Warehouse*. O modelo criado no BI Server foi utilizado para exemplificar essa tecnologia. Quando se modela uma fonte de dados para o Pentaho BI Server, automaticamente esse modelo fica disponível para exibir um cubo OLAP.

Ao acessar o cubo OLAP é exibida uma tabela contendo todos os campos do modelo e uma das medidas que o sistema configura como padrão. Após selecionar os dados e aplicar os filtros necessários através de interfaces gráficas, os dados são apresentados como na Figura 22. Nessa figura podemos ver um cubo configurado para exibir os cursos independente a que filial pertencem e o total de boletos em aberto em cada curso.

		Measures
		Num Emitido
		Turma
Curso	Parcela	<input checked="" type="checkbox"/> All Turmas
<input type="checkbox"/> All Cursos	<input type="checkbox"/> All Parcelas	2.591
#null	<input type="checkbox"/> All Parcelas	2
Administração	<input type="checkbox"/> All Parcelas	356
Ciência da Computação	<input type="checkbox"/> All Parcelas	21
Ciências Biológicas	<input type="checkbox"/> All Parcelas	29
Ciências Contábeis	<input type="checkbox"/> All Parcelas	99
Direito	<input type="checkbox"/> All Parcelas	638
ENGENHARIA AMBIENTAL	<input type="checkbox"/> All Parcelas	7
Educação Primária	<input type="checkbox"/> All Parcelas	8
Educação Física	<input type="checkbox"/> All Parcelas	98
Educação Infantil	<input type="checkbox"/> All Parcelas	107
Enfermagem	<input type="checkbox"/> All Parcelas	74
Engenharia Civil	<input type="checkbox"/> All Parcelas	72
Engenharia Elétrica	<input type="checkbox"/> All Parcelas	15
Ensino Fundamental	<input type="checkbox"/> All Parcelas	566
Ensino Médio	<input type="checkbox"/> All Parcelas	186
Ensino Médio II	<input type="checkbox"/> All Parcelas	13
Logística	<input type="checkbox"/> All Parcelas	10
Nutrição	<input type="checkbox"/> All Parcelas	26
PSICOLOGIA	<input type="checkbox"/> All Parcelas	38
Pedagogia	<input type="checkbox"/> All Parcelas	83
Pedagogia - Unices	<input type="checkbox"/> All Parcelas	39
Petróleo e Gas	<input type="checkbox"/> All Parcelas	31
Serviço Social	<input type="checkbox"/> All Parcelas	27
Sistemas de Informação	<input type="checkbox"/> All Parcelas	17
Tecnólogo Petróleo e Gas	<input type="checkbox"/> All Parcelas	24
Tecnólogo em Redes de Computadores	<input type="checkbox"/> All Parcelas	5

Figura 22: (Pentaho BI Server) Cubo OLAP - Boletos em aberto por Curso

Os cubos permitem que os dados sejam analisados sob diferentes perspectivas. Os mesmos dados filtrados acima podem ser apresentados como na Figura 23, apenas selecionando o campo curso para ser exibido nas linhas e as unidades nas colunas. Os dados de uma das colunas exibidas podem se transformar em colunas e revelar outros dados exibindo relacionamentos individuais a cada linha de resultado. Nessa figura pode-se ver o número de boletos em aberto por curso em qual filial eles ocorrem.

Curso	Filial													
	All Filiais	Carangola	Caratinga	Cataguases	Guarapari	Inhapim	Juiz de Fora	Leopoldina	Manhuaçu	Pingo de Gente	Serra	Teófilo Otoni	Vila Velha	Vitória
All Cursos	2.591	28	341	19	185	26	30	32	21	57	665	379	273	535
#null	2				2									
Administração	356							16			133	40	38	129
Ciência da Computação	21		21											
Ciências Biológicas	29									29				
Ciências Contábeis	99		39									60		
Direito	638	28	99		44		30	16	21		54			185
ENGENHARIA AMBIENTAL	7												7	
Educação Primária	8									8				
Educação Física	98				10									
Educação Infantil	107				16	1				49	25		10	6
Enfermagem	74		7	9							50	8		
Engenharia Civil	72		44									28		
Engenharia Elétrica	15		15											
Ensino Fundamental	566		72		81	25					160		113	115
Ensino Médio	186		16		24						20		71	55
Ensino Médio II	13		13											
Logística	10													10
Nutrição	26				8								18	
PSICOLOGIA	38											38		
Pedagogia	83									77				6
Pedagogia - Unices	39													39
Petróleo e Gas	31													31
Serviço Social	27		15										12	
Sistemas de Informação	17			10									7	
Tecnólogo Petróleo e Gas	24										24			
Tecnólogo em Redes de Computadores	5										5			

Figura 23: (Pentaho BI Server) Cubo OLAP - Distribuição por Curso relacionadas em cada Filial

O cubo OLAP é uma ferramenta poderosa e ao mesmo tempo de fácil utilização. Um usuário com pouca experiência em obtenção de informações de banco de dados não terá problemas em sua utilização após um pequeno treinamento, pois aplicando-se os filtros corretos é possível extrair um grande número de informações sem a necessidade de se criar novos relatórios.

### 3.4.1.3 Painéis Gerenciais

Os painéis gerenciais, ou painéis de controle são a representação gráfica do estado atual da dimensão do negócio analisada. Eles representam de forma visual e de fácil interpretação os resultados obtidos.

O Pentaho BI Server possui uma interface gráfica para criação dos painéis. Para criá-los é necessário um pouco mais de conhecimento em consultas SQL e desenvolvimento HTML. Mas isso não faz com que ele deixe de ser uma excelente alternativa na análise de dados, e tomada de decisões estratégicas. Para criá-lo é utilizada uma interface amigável para deixar o trabalho mais fácil.

A criação dos painéis passa pela elaboração de três elementos: o layout, os componentes e as fontes de dados. Esses elementos devem ser combinados para gerar os relatórios.

- **Layout:** Como o próprio nome indica, é um esboço, é um mapeamento de como os dados estarão dispostos para visualização. Conforme pode-se ver na Figura 24, nele são incluídos: linhas, colunas, espaços e adicionados arquivos de estilos. O layout segue o padrão HTML e aceita toda a formatação aplicada para tal. Cada elemento adicionado possui um nome e pode ter atributos como altura e largura. Os nomes são necessários para vincular outros componentes a esse design criado. Também podem ser adicionados modelos pré-definidos pelo sistema, agilizando o processo de criação dos painéis de controle.

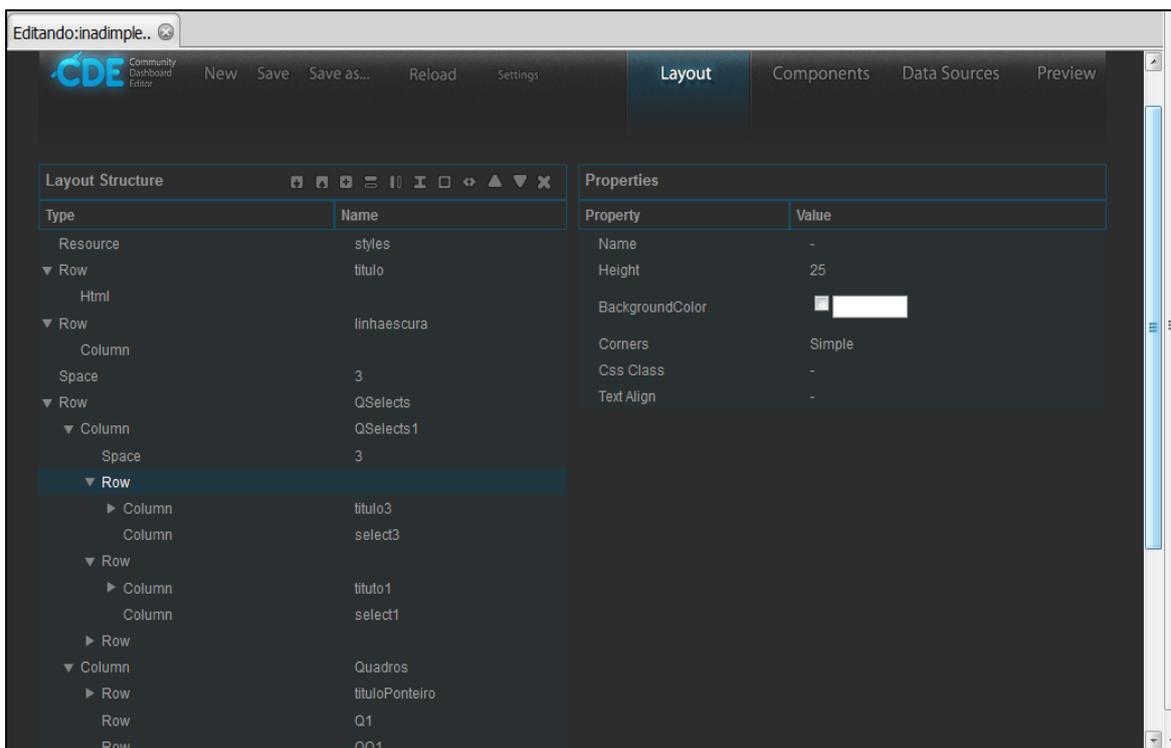


Figura 24: (Pentaho BI Server) Painel Gerencial – Criação de Layout para exibição.

- **Fontes de Dados:** As fontes de dados são os elementos responsáveis por coletar as informações nos bancos de dados e retornar os mesmos para ser aplicados aos elementos. O sistema aceita vários tipos de fontes de dados, como tabelas, arquivos de texto e vários formatos de bancos de dados.

As fontes de dados podem receber valores passados por outros elementos, ou alimentar outros elementos. Na Figura 25 tem-se os detalhes de uma fonte de dados adicionada. Ela faz conexão com o *Data Warehouse* boletos. É uma fonte de dados que recebe uma conexão e possui como atributo uma consulta SQL.

Os dados retornados pela consulta servirão para alimentar outros elementos. Esses elementos como será mostrado a seguir podem ser desde seletores até gráficos, mostradores e tabelas.

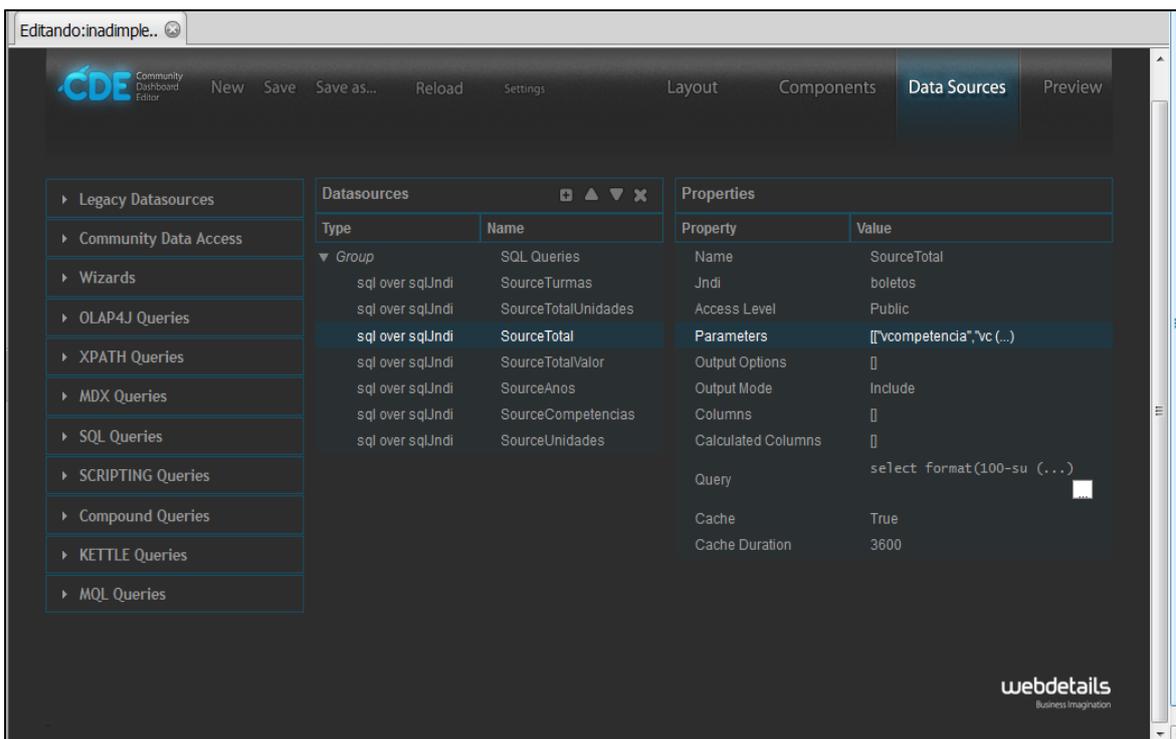


Figura 25: (Pentaho BI Server) Painel Gerencial – Criação de Fontes de Dados

- **Componentes:** Os componentes são os elementos visuais mais interessantes no painel de controle. Eles são os elementos utilizados para exibir os resultados obtidos através das fontes de dados adicionadas.

Os componentes podem ser elementos gráficos, tabelas, seletores e formulários. Um componente pode ser ligado a um objeto do layout e a uma fonte de dados. Além dos gráficos convencionais, ele dispõe de mostradores analógicos, que realçam o resultado obtido comparando-o a uma meta previamente estabelecida.

Cada componente é ligado a uma fonte de dados para alimentá-lo e um elemento do layout onde será exibido. A Figura 26 ilustra a interface para se trabalhar com componentes.

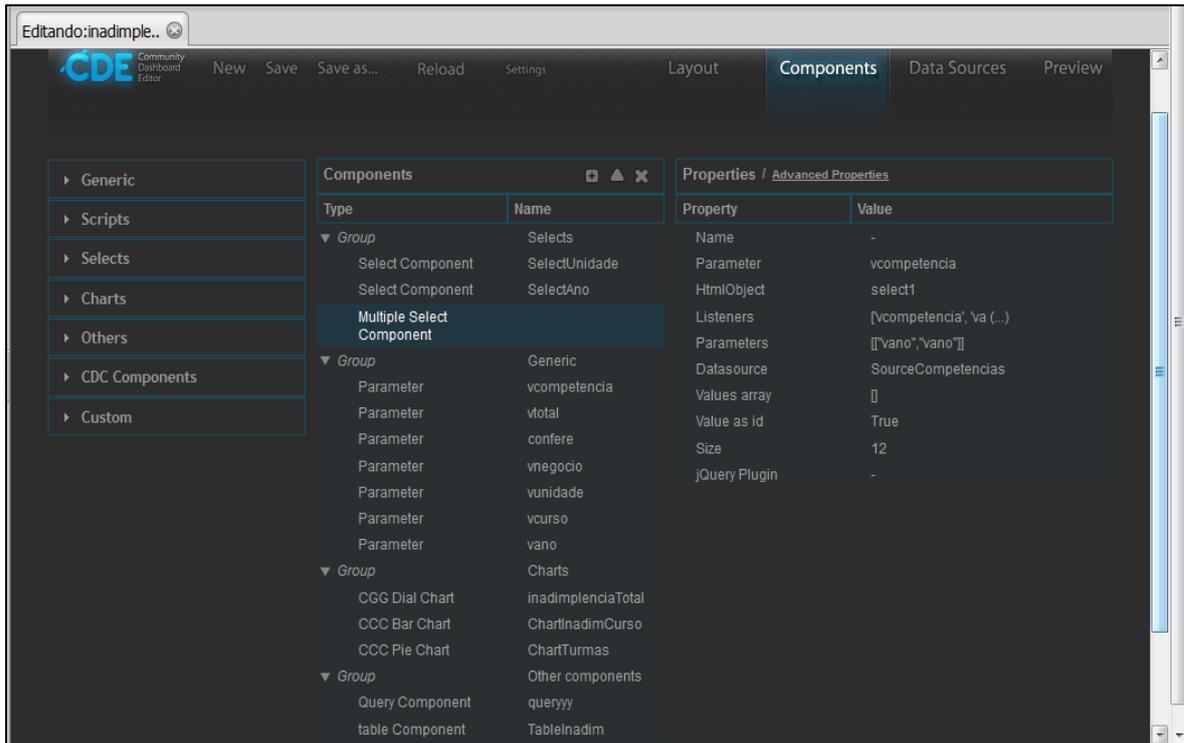


Figura 26: (Pentaho BI Server) Painel Gerencial: Criação de Componentes.

A união desses elementos exibirá um painel de controle conforme exibido na Figura 27. Esse painel poderá conter alguns seletores, como menu *dropdown* para tornar a informação mais rica, podendo ser selecionada em tempo real enquanto estiver sendo executada.

Os painéis também podem ser salvos em pastas do servidor, podendo ser acessado sempre que necessário. Ele trará sempre informações atualizadas, à medida que o repositório for atualizado.

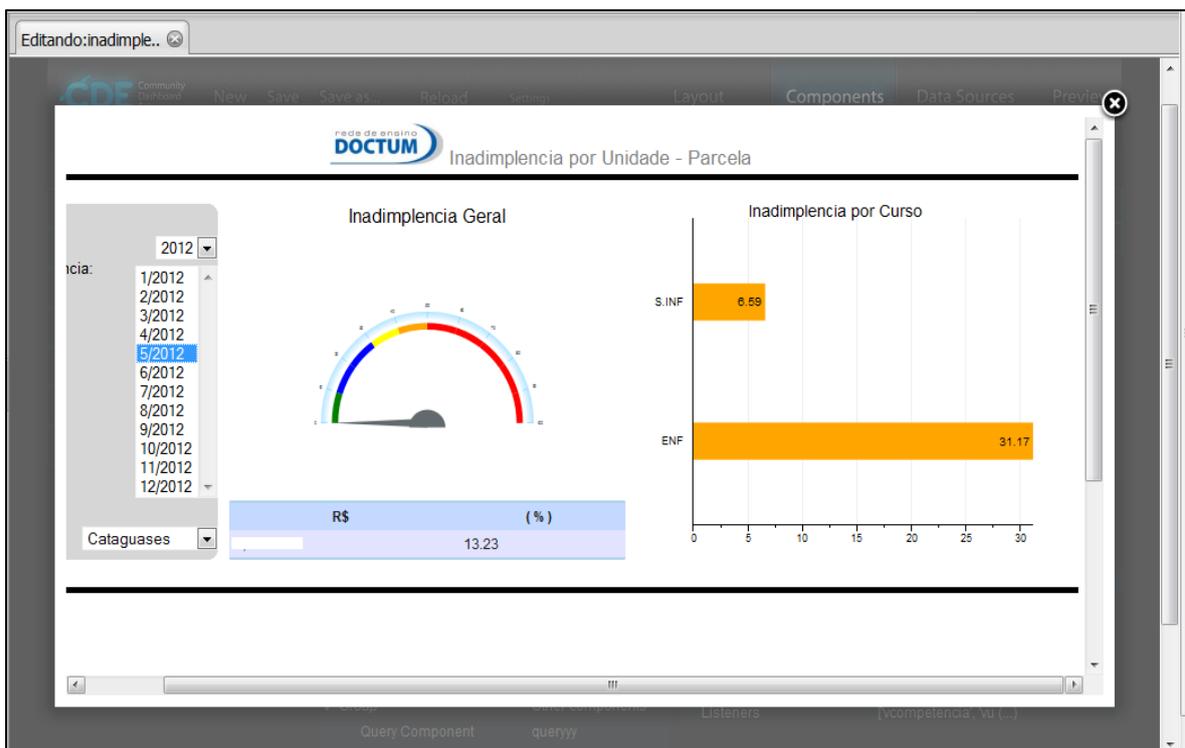


Figura 27: (Pentaho BI Server) Painel Gerencial – Visualização do Painel Gerencial Pronto.

Com esses três tipos de análises é possível atender o nível gerencial de uma empresa, utilizando recursos visuais diversificados e assim atender a todas as demandas exigidas.

## 4 RESULTADOS OBTIDOS

Nesse capítulo serão apresentados os resultados obtidos no processo de *Business Intelligence* aplicado nesse trabalho. Os resultados se encontram na forma de análises de informações que podem ser utilizadas para tomada de decisões.

Através do processo de extração, transformação e carga de dados utilizando-se o Kettle, gerou-se um *Data Warehouse* com dados de todas as unidades educacionais sumarizados e categorizados por filial, unidade de negócio, cursos, períodos e turmas, podendo chegar ao nível de alunos. Assim é possível um detalhamento gradual da informação até chegar ao nível mais detalhado, que é o aluno.

Com o Pentaho BI Server foi realizada análise das informações obtidas, demonstrando as principais funcionalidades da ferramenta. Os resultados se deram em forma de relatório, análises em um cubo OLAP e apresentação de painéis gerenciais.

### 4.1 RELATÓRIOS

Nessa seção será apresentado um relatório elaborado para apresentação de resultados consolidados das filiais. Primeiramente será apresentado um relatório de número de boletos gerados e o número de boletos pagos agrupados por curso com quebra por filial. Na Figura 28 pode-se observar o relatório, onde se vê uma quebra por filial, e um agrupamento por curso.

Trata-se de um relatório consolidado, apresentando os resultados de todas as filiais em um único relatório. Além disso, o usuário pode escolher quais colunas deseja visualizar, ou qual filtro usar. A interface permite que os modelos sejam agrupados, selecionados e filtrados da forma que o usuário precisar. Mesmo após salvo, o relatório pode ser alterado, acrescentando-se campos ou filtros. Esse é um exemplo de relatório consolidado, onde pode-se analisar os dados de todas as filiais em apenas um local, e que não é possível no AdX.

Todos os cursos dentro de suas respectivas filiais possuem totais de boletos emitidos e pagos, facilitando uma análise dos pagamentos desde um curso individualmente, até o total da rede.

RELATÓRIO DE BOLETOS EMITIDOS PARA COMPETENCIA 2012/05

Maio 14, 2012 @ 10:01

**Filial: Carangola**

Curso	Parcela	Num Emitidos	Num Pagos
Direito	5	282	205
<b>Total Carangola</b>		<b>282</b>	<b>205</b>

**Filial: Caratinga**

Curso	Parcela	Num Emitidos	Num Pagos
Ciência da Computação	5	69	61
Ciências Contábeis	5	183	164
Direito	5	576	498
Engenharia Civil	5	397	330
Engenharia Elétrica	5	107	91
Serviço Social	5	76	71
<b>Total Caratinga</b>		<b>1408</b>	<b>1215</b>

**Filial: Cataguases**

Curso	Parcela	Num Emitidos	Num Pagos
Enfermagem	5	41	35
Sistemas de Informação	5	88	82
<b>Total Cataguases</b>		<b>129</b>	<b>117</b>

Figura 28: (Pentaho BI Server) Relatório de número de Boletos por Curso com quebra por Filial

Na Figura 29 observa-se o mesmo relatório agrupado por filiais e com quebra por curso. Ele mostra como o mesmo modelo pode ser utilizado de várias formas. Nesse relatório podemos comparar os números de um curso em filiais diferentes.

Ele permite que seja analisado em quais unidades existe determinado curso. Também pode ser analisado o número total de boletos emitidos para cada curso, permitindo saber qual curso com maior número de alunos na rede, ou que se calcule o percentual de inadimplência médio de cada curso.

RELATÓRIO		Maio 23, 2012 @ 10:38	
<b>RELATÓRIO DE BOLETOS EMITIDOS PARA COMPETENCIA 2012/05</b>			
<b>Curso: Administração</b>			
Filial	Parcela	Num Emitidos	Num Pagos
Guarapari	5	50	47
Leopoldina	5	145	132
Serra	5	470	400
Teófilo Otoni	5	115	91
Vila Velha	5	113	97
Vitória	5	552	459
<b>Total Administração</b>		<b>1445</b>	<b>1226</b>
<b>Curso: Ciência da Computação</b>			
Filial	Parcela	Num Emitidos	Num Pagos
Caratinga	5	69	61
<b>Total Ciência da Computação</b>		<b>69</b>	<b>61</b>
<b>Curso: Ciências Biológicas</b>			
Filial	Parcela	Num Emitidos	Num Pagos
Serra	5	99	89
<b>Total Ciências Biológicas</b>		<b>99</b>	<b>89</b>
<b>Curso: Ciências Contábeis</b>			
Filial	Parcela	Num Emitidos	Num Pagos
Caratinga	5	183	164
Guarapari	5	10	10
Teófilo Otoni	5	162	136
<b>Total Ciências Contábeis</b>		<b>355</b>	<b>310</b>

Figura 29: (Pentaho BI Server) Relatório de Boletos por Filiais com quebra por Curso

Com isso demonstra-se a flexibilidade que o usuário possui em configurar seus relatórios alterando o foco de sua análise através da criação de novos relatórios ou alteração nos relatórios já existentes.

Atualmente esses relatórios consolidados são gerados coletando-se os dados de cada instância do AdX e criando-se manualmente um relatório em uma planilha eletrônica. Com os relatórios do Pentaho BI Server pode-se também alterar o foco e ver comparações diferentes em um mesmo modelo de dados.

## 4.2 CUBO OLAP

O cubo OLAP é o método de análise mais explorado em *Business Intelligence*. A rapidez e a facilidade com que se pode alterar o foco de visualização da informação contribui com esse fato (MACHADO, 2007). Isso permite que várias análises sejam realizadas em um mesmo modelo de dados,

apenas alterando a forma como os dados são apresentados, agrupando dados de diferentes maneiras e combinando-os da maneira que permita a análise necessária.

Nos resultados apresentados foi utilizado o modelo de receitas, filtrando os boletos em aberto no primeiro semestre do ano de dois mil e doze. Como esse filtro aplicado nesse modelo foi possível obter várias análises, fazendo comparações de cursos por unidade, cursos por parcela, cursos com quebra por turma e separadas por parcela, ou separadas por unidade, além de exibir um gráfico de distribuição de boletos em aberto por filial.

Na Figura 30 pode-se ver um cubo gerado totalizando o número de boletos em aberto por curso. Esse cubo pode ser exibido de forma diferente mostrando, por exemplo, a distribuição de boletos em aberto por curso e em cada parcela.

		Measures
		Num Emitido
		Turma
Curso	Parcela	All Turmas
All Cursos	All Parcelas	2.591
#null	All Parcelas	2
Administração	All Parcelas	356
Ciência da Computação	All Parcelas	21
Ciências Biológicas	All Parcelas	29
Ciências Contábeis	All Parcelas	99
Direito	All Parcelas	638
ENGENHARIA AMBIENTAL	All Parcelas	7
Educacao Primaria	All Parcelas	8
Educação Física	All Parcelas	98
Educação Infantil	All Parcelas	107
Enfermagem	All Parcelas	74
Engenharia Civil	All Parcelas	72
Engenharia Elétrica	All Parcelas	15
Ensino Fundamental	All Parcelas	566
Ensino Médio	All Parcelas	186
Ensino Médio II	All Parcelas	13
Logística	All Parcelas	10
Nutrição	All Parcelas	26
PSICOLOGIA	All Parcelas	38
Pedagogia	All Parcelas	83
Pedagogia - Unices	All Parcelas	39
Petróleo e Gas	All Parcelas	31
Serviço Social	All Parcelas	27
Sistemas de Informação	All Parcelas	17
Tecnólogo Petroleo e Gas	All Parcelas	24
Tecnólogo em Redes de Computadores	All Parcelas	5

Figura 30: (Pentaho BI Server) Numero de Boletos em aberto por Curso

Na Figura 31 pode-se ver o cubo aberto por parcelas. Isso se torna útil para realizar comparações, analisar tendências e obter respostas sobre determinada perspectiva de maneira rápida.

Vale ressaltar que apenas foi alterado a disposição dos dados, exibindo as parcelas como colunas. Isso permite analisar os cursos parcela a parcela. O cubo permite que se tenha uma visão de qualquer ponto de vista, desde que ele esteja inserido no modelo dimensional. Nessa análise pode-se ver o número de boletos em aberto de cada curso por parcela.

Curso	Parcela						
	All Parcelas	1	2	3	4	5	6
	Measures						
	● Num Emitido						
All Cursos	2.591	140	398	446	505	528	574
#null	2	1	1				
Administração	356		53	60	74	81	88
Ciência da Computação	21		4	4	4	4	5
Ciências Biológicas	29		3	5	6	8	7
Ciências Contábeis	99		17	20	21	20	21
Direito	638	2	99	124	128	137	148
ENGENHARIA AMBIENTAL	7		2	1	1	1	2
Educação Primária	8			1	2	2	3
Educação Física	98		17	19	20	20	22
Educação Infantil	107	12	14	14	20	22	25
Enfermagem	74	1	12	10	14	19	18
Engenharia Civil	72		11	12	15	16	18
Engenharia Elétrica	15		3	3	3	3	3
Ensino Fundamental	566	88	88	92	97	97	104
Ensino Médio	186	33	24	27	36	32	34
Ensino Médio II	13	2	2	2	2	2	3
Logística	10		1	2	2	2	3
Nutrição	26		4	4	6	6	6
PSICOLOGIA	38		7	8	8	8	7
Pedagogia	83		12	14	18	18	21
Pedagogia - Unices	39		7	8	8	8	8
Petróleo e Gas	31		5	6	6	6	8
Serviço Social	27		5	5	5	6	6
Sistemas de Informação	17		4	2	3	3	5
Tecnólogo Petróleo e Gas	24	1	2	2	5	6	8
Tecnólogo em Redes de Computadores	5		1	1	1	1	1

Slicer: [Pagamento=#null] [Periodo Letivo=2011/02]

Figura 31: (Pentaho BI Server) Cubo OLAP - Custos x Parcelas

A análise compara os dados presentes no modelo, levando-se em consideração os filtros aplicados e as dimensões selecionadas para serem exibidas. Na Figura 32 pode-se observar o mesmo modelo sendo visualizado por filial.

Nessa visualização é possível analisar que filiais compartilham os mesmos cursos com parcelas sem pagamento. Com uma seleção é possível alterar o foco da análise, revelando novas informações.

Curso	Measures													
	Num Emitido													
	Filial													
	All Filiais	Carangola	Caratinga	Cataguases	Guarapari	Inhapim	Juiz de Fora	Leopoldina	Manhuaçu	Pingo de Gente	Serra	Tedfio Ottoni	Vila Velha	Vitória
All Cursos	2.591	28	341	19	185	26	30	32	21	57	665	379	273	535
#null	2				2									
Administração	356							16			133	40	38	129
Ciência da Computação	21		21											
Ciências Biológicas	29									29				
Ciências Contábeis	99		39									60		
Direito	638	28	99		44		30	16	21		54			185
ENGENHARIA AMBIENTAL	7											7		
Educação Primária	8									8				
Educação Física	98				10									
Educação Infantil	107				16	1				49	25		10	6
Enfermagem	74		7	9								8		
Engenharia Civil	72		44							50		28		
Engenharia Elétrica	15		15											
Ensino Fundamental	566		72		81	25				160			113	115
Ensino Médio	186		16		24					20			71	55
Ensino Médio II	13		13											
Logística	10													10
Nutrição	26				8							18		
PSICOLOGIA	38											38		
Pedagogia	63									77				6
Pedagogia - Unices	39													39
Petróleo e Gas	31													31
Serviço Social	27		15									12		
Sistemas de Informação	17			10								7		
Tecnólogo Petróleo e Gas	24									24				
Tecnólogo em Redes de Computadores	5										5			

Figura 32: (Pentaho BI Server) Cubo OLAP - Cursos x Filiais

Se for preciso uma análise no próximo nível, por turma, basta acrescentar a coluna e expandir os cursos, para poder obter os resultados de turmas, agrupados em seus respectivos cursos, e abertas por filial. A Figura 33 mostra essa visualização. Caso fosse necessário analisar as turmas, sem quebra por curso, seria possível, apenas adicionando a coluna turma antes da coluna curso.

Curso	Turma	Measures													
		Num Emitido													
		Filial													
		All Filiais	Carangola	Caratinga	Cataguases	Guarapari	Inhapim	Juiz de Fora	Leopoldina	Manhuaçu	Pingo de Gente	Serra	Tedfio Ottoni	Vila Velha	Vitória
Direito	DIR6NB	3				1									
	DIR7N	13	8		5										
	DIR8M	15			4					5					6
	DIR8N	10					5								
	DIR8NA	5			5										
	DIR8NB	7													7
	Mod_JED II	1									1				
ENGENHARIA AMBIENTAL	All Turmas	7											7		
	ENG.AMB2N	7											7		
Educação Primária	All Turmas	8									8				
	V-1ºAno	4									4				
	V-2ºAno	3									3				
	V-3ºAno	1									1				
Educação Física	ED.FIS1N	31										31			
	ED.FIS2NA	10										10			
	ED.FIS2NB	13										13			
	ED.FIS4NA	10										10			
	ED.FIS4NB	4										4			
	ED.FIS6N	20										20			
Educação Física	All Turmas	98				10									
	Efísica8N	10				10									
Educação Infantil	All Turmas	107				16	1			49	25		10	6	
	EI-NIV.I-VES	16				3					13				
	EI-NIV.I-VESB	2											2		
	EI-NIV.II-VES	8				7								1	
	EI-NIV.II-VESA	6									5		1		
	EI-NIV.II-VESB	1									1				
	EI-NIV.III-VES	11				6								5	
	EI-NIV.III-VESA	6									6				
	EI-NIV.III-VESB	7												7	
	I-2ºPeríodo	11									11				
	I-Maternal I	2										2			
	I-Maternal II	6										6			
	MATERNAL	1						1							
V-1ºPeríodo	6									6					
V-2ºPeríodo	2										2				

Figura 33: (Pentaho BI Server) Cubo OLAP - Cursos e Turmas x Filial

Abaixo são mostradas outras formas de visualização. Na Figura 34 pode-se ver os cursos por linha, com quebras por parcelas, e as turmas abertas em colunas. Isso mostra como o cubo OLAP permite uma vasta combinação de seus dados.

Turma														
Parcela	Curso	Measures												
		#null	1º Ano	1º ANO	2º Ano	2º ANO	3º ANO	4º Ano	5ª SERIE	6ª SERIE	7ª SÉRIE	8ª SÉRIE	9º ANO	
		Measures												
		Num Emitido												
All Parcelas	All Cursos	2.591	2	1	7	6	9	13	6	6	12	20	34	12
1	All Cursos	140	1		1	1	2	2	1	1	2	3	6	2
	#null	1	1											
	Direito	2												
	Educação Infantil	12												
	Enfermagem	1												
	Ensino Fundamental	88			1	1			1	1	2	3	6	2
	Ensino Médio	33			1		2							
	Ensino Médio II	2						2						
	Tecnólogo Petróleo e Gas	1												
2	All Cursos	398	1		1	1	1	2	1	1	2	3	6	2
	#null	1	1											
	Administração	53												
	Ciência da Computação	4												
	Ciências Biológicas	3												
	Ciências Contábeis	17												
	Direito	99												
	ENGENHARIA AMBIENTAL	2												
	Educação Física	17												
	Educação Infantil	14												
	Enfermagem	12												
	Engenharia Civil	11												
	Engenharia Elétrica	3												
	Ensino Fundamental	88			1	1			1	1	2	3	6	2
	Ensino Médio	24			1		1							
	Ensino Médio II	2						2						
	Logística	1												
	Nutrição	4												
	PSICOLOGIA	7												
	Pedagogia	12												
	Pedagogia - Unices	7												
	Petróleo e Gas	5												
	Serviço Social	5												
	Sistemas de Informação	4												
	Tecnólogo Petróleo e Gas	2												
	Tecnólogo em Redes de Computadores	1												
3	All Cursos	444			1	1	1	1	1	1	1	1	1	1

Figura 34: (Pentaho BI Server) Cubo OLAP – Cursos agrupados por Parcelas x Turmas

Na Figura 35 pode-se ver um gráfico gerado pelo Pentaho BI Server, com base nos dados exibidos no cubo. Isso torna a análise mais amigável, pois os gráficos permitem uma melhor interpretação do quadro avaliado.

O Pentaho BI Server permite a exibição de vários tipos de gráficos, além de exportar os dados exibidos para planilhas eletrônicas e permitir a impressão dos mesmos. Na Figura 35 podemos ver os ícones de planilha eletrônica e impressora, auxiliando nessas tarefas.

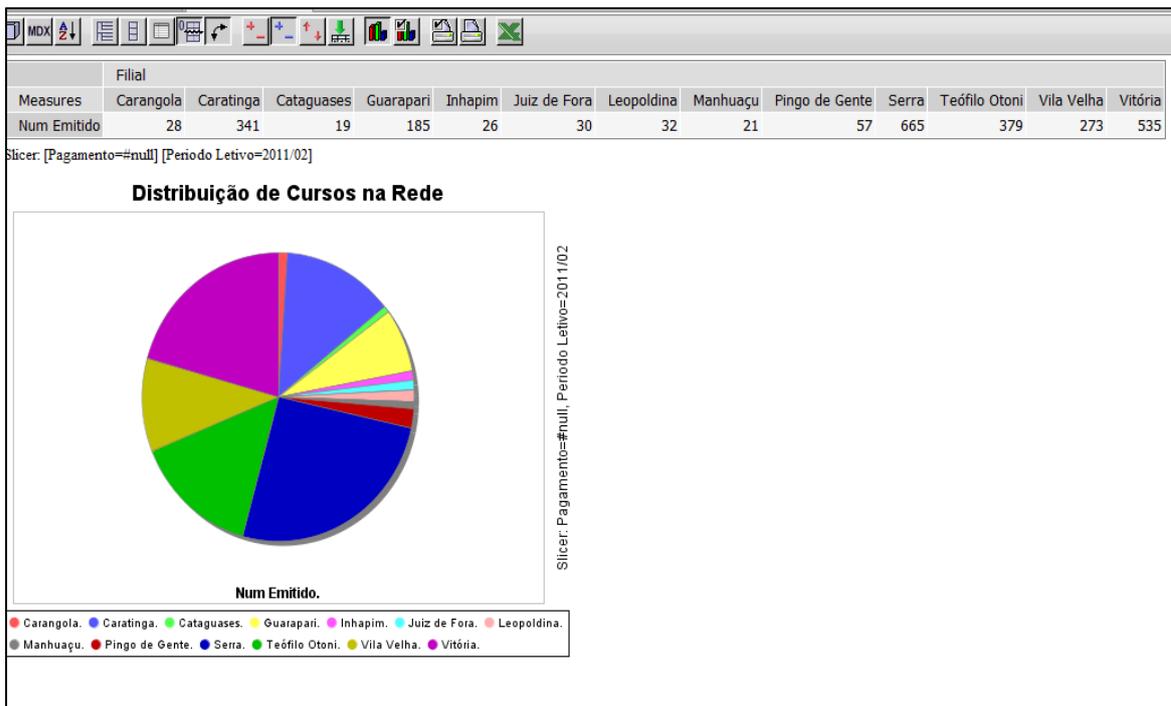


Figura 35: (Pentaho BI Server) Cubos OLAP - Gráfico gerado na interface geradora de Cubo OLAP

### 4.3 PAINÉIS GERENCIAIS

Painéis gerenciais são instrumentos de análise utilizados por tomadores de decisão para análise de informações através de gráficos, mostradores, tabelas e semáforos, que fazem o papel de indicadores de desempenho e facilitando a visualização do resultado esperado (SCHEPS, 2008).

Os painéis gerenciais do Pentaho BI Server fornecem vários indicadores para melhor análise das informações. Para demonstrar os resultados nesse trabalho foram criados painéis para medir a inadimplência das unidades.

O primeiro painel mostra os indicadores para acompanhamento de inadimplência de todas as filiais. Na Figura 37 pode-se ver esse painel. No lado superior esquerdo vemos os seletores.

Ao se selecionar um ano, o seletor de competência é carregado com os meses (competências) daquele ano. O terceiro seletor se refere à unidade de negócio a ser analisada: graduação, colégio, pós-graduação, etc.

Na parte superior central há um mostrador analógico que mostra a inadimplência geral de todas as filiais. Esse mostrador é auxiliado por cores indicativas de posição que refletem a posição apontada pelo mostrador, realçando o resultado obtido

Abaixo do mostrador é exibida uma tabela com valores importantes a serem analisados, como valor emitido, valor recebido, valor de descontos concedidos, valor de inadimplência e o percentual apontado pelo mostrador analógico. Esses dados ajudam na interpretação do mostrador, enriquecendo a análise.

Na parte superior direita é exibido um gráfico em ordem decrescente, classificando as unidades pelo percentual de inadimplência. Ele mostra o percentual de inadimplência de cada unidade. Na inferior do leiaute é feita uma classificação por filial, utilizando o valor da inadimplência de cada unidade. Essa classificação pode ser usada para comparar percentual e valor de inadimplência por filial.

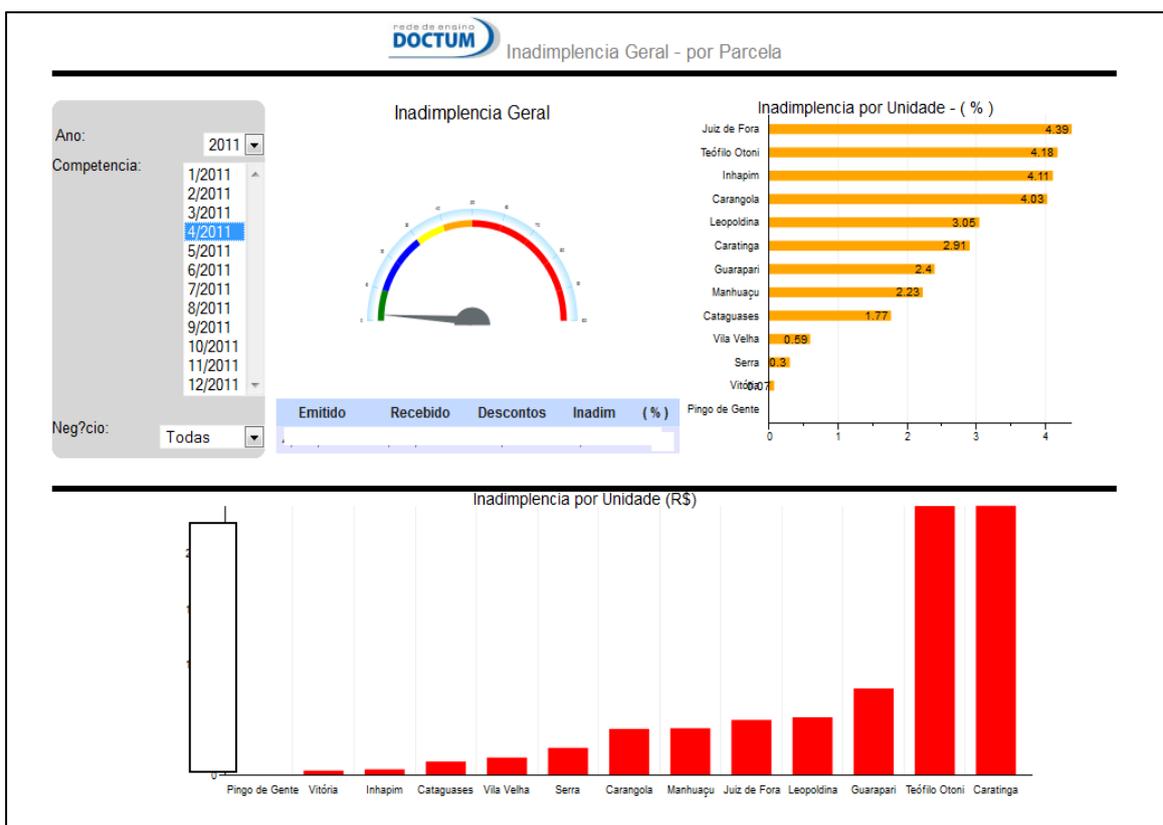


Figura 36: (Pentaho BI Server) Painel Gerencial - Inadimplência Geral

Esse gráfico auxilia na análise total e por unidade. Para analisar dados específicos por unidade foi criado outro painel. Para ter acesso a esse painel basta clicar sobre uma das barras do gráfico de inadimplência percentual por unidade. Ao clicar em uma das barras o Pentaho BI Server outro painel de inadimplência por unidade, mostrado na Figura 37.

O painel de acompanhamento de inadimplência por filial faz uma análise mais profunda da filial. Na direita podemos ver os seletores, para filtrar as competências e a filial que se deseja analisar.

Na área central superior foi inserido um mostrador analógico, que indica em que faixa a inadimplência se encontra. Ele dá uma posição geral da filial, fácil de ser analisado e com cores que auxiliam sua interpretação.

No lado esquerdo pode-se ver um gráfico indicador de inadimplência por curso. Ele também reflete os filtros utilizados, e mostra os indicadores de inadimplência de cada curso.

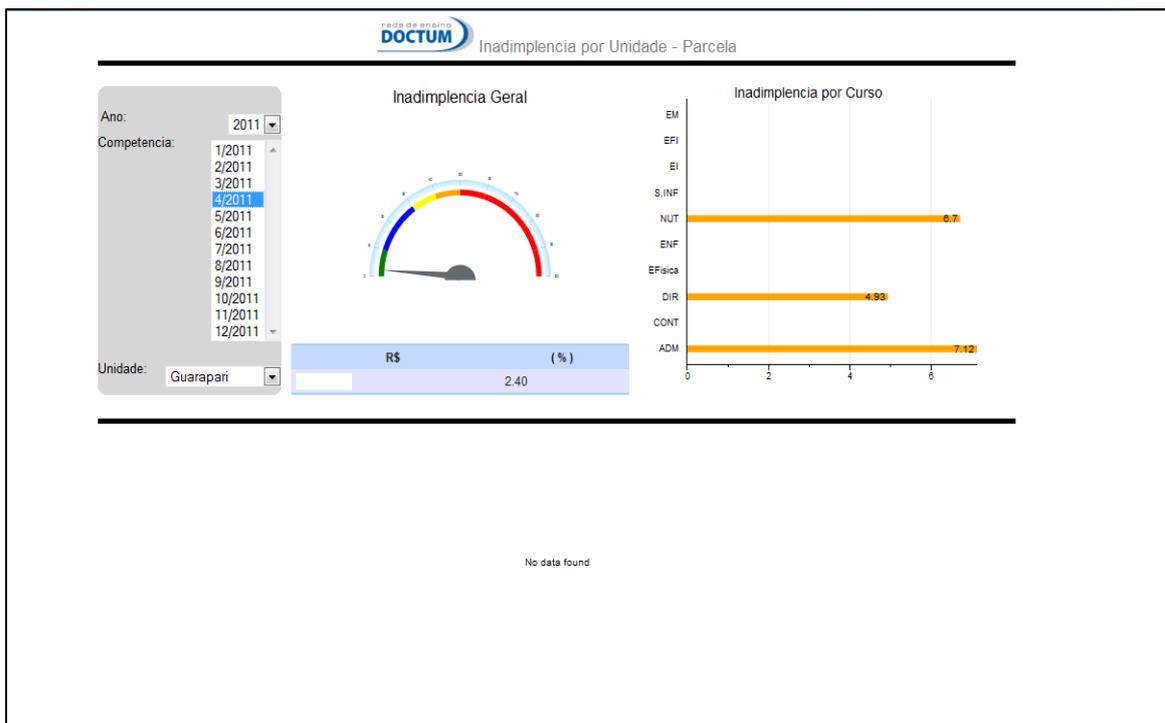


Figura 37: (Pentaho BI Server) Painel Gerencial - Inadimplência por Filial

O painel Figura 37 possui mais um recurso. Ao se clicar na barra exibida para cada curso no gráfico de inadimplência por curso, é exibido na parte inferior

um gráfico de pizza com a distribuição de inadimplência por turma, referente ao curso selecionado.

Esse gráfico por turma pode ser observado na Figura 38. Nele pode-se ver a distribuição de inadimplência nas turmas do curso de direito da filial de Guarapari na competência 04/2011.

Para analisar a inadimplência de turmas de outro curso, basta clicar na barra correspondente ao resultado desse curso para que o gráfico por turma recarregue, assim como se queira analisar outra competência ou mais de uma competência, basta selecionar essa competência.

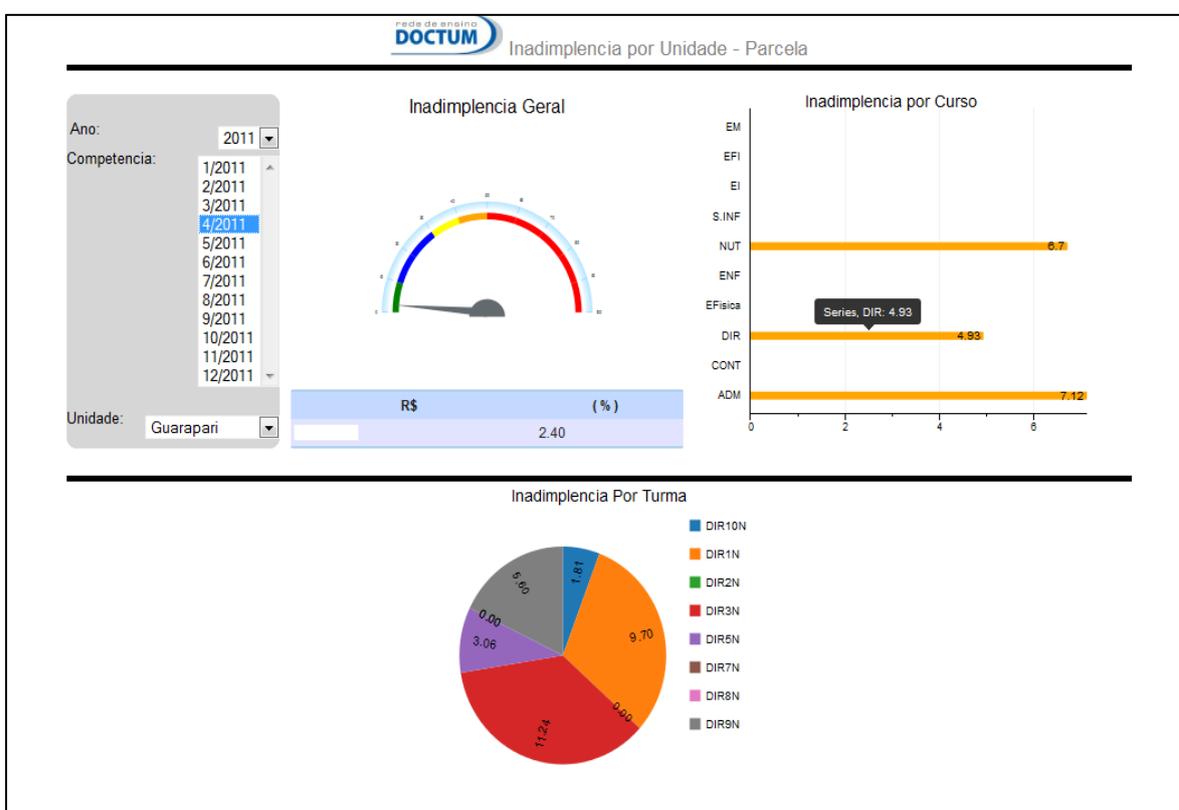


Figura 38: (Pentaho BI Server) Painéis Gerenciais - Distribuição por Turma

Os painéis tornam a análise mais intuitiva, pois visualmente é mais fácil perceber, por exemplo, qual dos cursos possui maior inadimplência, ou perceber se há relação entre o percentual de inadimplência das filiais e o valor de inadimplência. Com essa análise intuitiva é possível ir navegando pelas regionais, filiais, curso e turmas, descobrindo quais turmas são as maiores responsáveis pela inadimplência.

## 5 CONCLUSÃO

A utilização de técnicas de *Business Intelligence* mostrou-se um forte aliado na busca de resultados consolidados para realizar a análise financeira da rede de ensino Doctum, através da obtenção dos dados de diversos bancos de dados em um único local e a utilização dos mesmos na elaboração de relatórios OLAP.

O processo aplicado atualmente para a obtenção de informações consolidadas de toda a instituição prevê a geração de relatórios em cada instância do AdX. Desse conjunto de relatórios são extraídas informações que alimentam planilhas que finalizam o processo de consolidação dos dados. Esse processo além de demorado é suscetível a erros, pois na transcrição de dados de um relatório para uma planilha podem ser digitados valores errados, ou mesmo em locais errados, descaracterizando a informação gerada. Assim a tomada de decisões é realizada sobre dados que podem não retratar a realidade presenciada pela empresa.

O processo de ETL centraliza os dados de vários bancos de dados, permitindo que sejam criados relatórios com maior rapidez. Essa centralização dos dados é obtida em menos de 1 hora e diariamente os dados são atualizados através de rotina automatizada. Levando-se em consideração que a consolidação manual dos dados visa à elaboração de apenas relatórios específicos, com o processo de ETL tem-se um ganho significativo, pois após a criação do DW os dados modelados tornam-se fonte de dados para diversos relatórios. A precisão dos dados gerados também pode ser vista com um ganho significativo, pois como o processo de ETL é automático ele está menos propício a erros, uma vez que não há intervenção humana na sua execução.

Os relatórios fornecidos pelo Pentaho BI Server permitem uma melhor visualização da informação gerada. A utilização dos recursos OLAP presentes nele são capazes de exibir relatórios consolidados ou separadamente por filiais, unidades de negócio, cursos, turmas, agrupando os dados da forma que for necessária. Esses relatórios dão liberdade para que o tomador de decisão elabore as análises necessárias para realizar a tomada de decisões.

A visualização dos cubos OLAP permite que o usuário mude o foco de visão dos dados analisados, procurando tendências e realizando comparações nos dados revelados. Os cubos OLAP permitem que se escolha quais campos gerados no modelo de dados serão exibidas e também que os dados exibidos em determinado campo se tornem colunas, revelando novos dados.

Os relatórios permitem gerar informações consolidadas, agrupadas por qualquer campo do modelo de dados, e aplicando-se vários filtros aos mesmos para se chegar à informação buscada. Eles são úteis na obtenção de relatórios padronizados e configurados para impressão.

Os painéis gerenciais possibilitam uma interação entre seus diversos elementos, sendo eles gráficos de diversos formatos, tabelas, mostradores analógicos, e seletores. Eles permitem que intuitivamente o usuário navegue pelos dados exibidos, clicando nos seletores, gráficos e mesmo nas tabelas, exibindo novos dados filtrados pela fatia de dados que representam. Ou seja, se existe um gráfico de barras é possível configurá-lo para que um clique em cada barra exiba informações sobre os dados que a barra representa. Isso melhora o entendimento do usuário, pois os elementos gráficos ajudam na interpretação dos dados exibidos e essa interação faz com que ele crie uma linha de raciocínio lógica.

As ferramentas de BI possibilitaram a criação de mecanismos que permitem a análise consolidada dos dados financeiros da rede de ensino Doctum, contribuindo para que a tomada de decisão possa ser realizada sobre informações obtidas de forma mais rápida e com menos possibilidade de erros. Os resultados obtidos mostram que *Business Intelligence* é aplicável na rede de ensino Doctum consolidando dados e dando suporte à tomada de decisões, e que a suíte Pentaho contribui para esse processo através de ferramentas bem elaboradas e com grande poder de criação de relatórios para auxiliar a análise de dados.

## 6 TRABALHOS FUTUROS

Como trabalhos futuros pode-se utilizar outras ferramentas da Suíte Pentaho, pois nesse trabalho o foco foi dado a apenas duas ferramentas, sendo elas o Pentaho Kettle e o Pentaho BI Server. Um exemplo de ferramenta a ser explorada é o Pentaho Reporter Designer. Trata-se de uma ferramenta utilizada para gerar modelos de relatórios, sendo publicados no Pentaho BI Server. Além de relatórios prontos, podem ser criados modelos com opções de filtros, dando mais utilidade aos relatórios.

Também pode ser trabalhado outras áreas do sistema AdX não utilizadas aqui. A análise dos dados acadêmicos de forma consolidada trará grandes benefícios para a instituição. Com *Business Intelligence* será possível criar mecanismos para comparar informações entre filiais, e assim poder obter análises mais consolidadas, mais precisas e de forma mais rápida da rede de ensino Doctum como um todo.

## REFERÊNCIAS

- AMBIENTE LIVRE. **Conhecendo o Pentaho BI Server - Arquitetura e Infra-estrutura**, 2012. Disponível em: <<http://www.ambientelivre.com.br/tutoriais-pentaho-bi/133-conhecendo-o-pentaho-bi-server-arquitetura-e-infra-estrutura.html>>. Acesso em: 22 maio. 2012.
- ANZANELLO, C. A. OLAP Conceitos e Utilização. **Universidade Federal do Rio Grande do Sul**, p. 7, 2008.
- COELHO, F. Z.; SOUZA, M. A. V. F. DE; BALDAM, R. DE L. Business Intelligence : suporte à gestão do desempenho científico. **cefetes.br**, p. 4, 2011.
- CRAMER, R. **Estudo analítico de ferramentas open source para Ambientes OLAP**. Criciúma: Universidade do Extremo Sul Catarinense, 2006. v. 1p. 65
- ELMARI, R.; NAVATHE, S. B. **SISTEMAS DE BANCO DE DADOS**. 4ª Edição ed. São Paulo: Pearson Education do Brasil, 2005. p. 724
- FLUX SOFTWARES, T. **AdX - Gerenciamento Acadêmico**, 2006. Disponível em: <<http://www.doctumtec.com.br:9673/doctumtec/produtos/desenvolvimento?abrir=adx>>. Acesso em: 24 maio. 2012.
- FORTULAN, M. R. UMA PROPOSTA DE APLICAÇÃO DE BUSINESS INTELLIGENCE NO CHÃO-DE-FÁBRICA. **GESTÃO & PRODUÇÃO**, v. 12, p. 12, 2005.
- FREITAS, M. **Data Warehouse - Conceitos**, 2010. Disponível em: <<http://www.modernopapo.com.br/?p=1567>>. Acesso em: 24 abr. 2012.
- FUJIWARA, Daniel Kenji. **Data Warehouse como Instrumento de Suporte à Avaliação Acadêmica**. Brasília: Universidade de Brasília, 2006.
- GOMES, A. **Tutorial Kettle - Pentaho Data Integration**, 2012. Disponível em: <<http://www.ademargomes.com/index.php/artigos/56-tutorialkettle.html>>. Acesso em: 22 maio. 2012.
- GUSMAN, J. **Pentaho BI Server**, 2012. Disponível em: <[http://wiki.bizcubed.com.au/xwiki/bin/view/Pentaho Tutorial/BI Server](http://wiki.bizcubed.com.au/xwiki/bin/view/Pentaho+Tutorial/BI+Server)>. Acesso em: 22 maio. 2012.
- HORITA, F. **Modelagem Dimensional**. Disponível em: <[http://www.flaviohorita.stillit.com.br/index.php?option=com\\_content&view=article](http://www.flaviohorita.stillit.com.br/index.php?option=com_content&view=article)>

&id=81:modelagem-dimENSIONAL&catid=43:business-intelligence&Itemid=57>.  
Acesso em: 7 jun. 2012.

KIMBALL, R.; ROSS, M. **The Complete Guide to Dimensional Modeling**. 2. ed. New York: Robert Ipsen, 2002. p. 449

MACHADO, F. N. R. **Tecnologia e Projeto de DataWarehouse**. 3. ed. São Paulo: Érica, 2007. p. 320

MOREIRA, E. **Modelo Dimensional para Data Warehouse**, 2006. Disponível em: <[http://imasters.com.br/artigo/3836/modelo\\_](http://imasters.com.br/artigo/3836/modelo_)>. Acesso em: 24 abr. 2012.

OLIVEIRA, Ana Júlia de. **ESTUDO DE DATA WAREHOUSE EM SISTEMA DE APOIO A DECISÕES**. Caratinga: FIC, 2006.

PENTAHO COMMUNITY. **Introducing the Pentaho BI Suite 3 . 5 .** . Orlando: [s.n.]. , 2009.

PENTAHO COMUNITY. **Pentaho Kettle Project**. Disponível em: <<http://kettle.pentaho.com/>>. Acesso em: 22 maio. 2012.

PENTAHO COMUNITY. **Pentaho BI Platform / Server**. Disponível em: <[http://community.pentaho.com/projects/bi\\_platform/](http://community.pentaho.com/projects/bi_platform/)>. Acesso em: 22 maio. 2012.

PENTAHO CORPORATION. **Pentaho Open Source Business Intelligence Platform: Technical White Paper**. . Orlando: [s.n.]. , 2006.

PENTAHO CORPORATION, T. **PENTAHO DATA INTEGRATION**. Disponível em: <<http://www.pentaho.com/explore/pentaho-data-integration/>>. Acesso em: 22 maio. 2012.

SCHEPS, S. **Business Intelligence for Dummies**. 1. ed. Indianapolis: Wiley Publish, 2008. p. 358

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **SISTEMA DE BANCO DE DADOS**. TERCEIRA E ed. São Paulo: Elsevier Editora Ltda, 2006. p. 808

SOFTWARE WIKI. **Reliable business intelligence : Datawarehouse, Datamart and Datamining**, 2012. Disponível em: <<http://swiki.net/reliable-business-intelligence-datawarehouse-datamart-and-datamining.html>>. Acesso em: 5 jun. 2012.

SOUZA, M. DE. **Ferramentas OLAP**, 2003. Disponível em: <[http://imasters.com.br/artigo/1498/gerencia/ferramentas\\_olap/](http://imasters.com.br/artigo/1498/gerencia/ferramentas_olap/)>. Acesso em: 21 maio. 2012.

RODRIGUES, D. **What is business intelligence - ETL**, 2012. Disponível em <<http://projetoseti.com.br/gestao/o-que-e-business-intelligence-parte-iv-v-etl/>>.

Acesso em 25 maio. 2012.