

Neuro BI: Uma Ferramenta *Web open source* para apoio à Tomada de Decisão

Abstract. *The use of technologies for data analysis favors the organizations visibility, it highlights information contained in a larger information sets. This paper presents an open source web tool which main goal is to allow queries creation in data cubes without the need for knowledge of the mdx language.*

Resumo. *A utilização de tecnologias para análise de dados favorece a visibilidade em qualquer organização, pois realça informações contidas em um conjunto de informações maiores. Este artigo descreve uma ferramenta web open source que tem como objetivo principal permitir a criação de consultas em cubos de dados sem a necessidade de conhecimento da linguagem mdx.*

1. Introdução

A utilização de ferramentas OLAP tem-se difundido nos últimos anos pela disponibilidade de soluções proprietárias e de código aberto (*open source*). Como soluções proprietárias destaca-se o *Microsoft Analysis Services*, distribuído com o *MS-SQL Server*. Entre as ferramentas de código aberto para implementação de ferramentas OLAP destaca-se o Pentaho, através da utilização da biblioteca *JPivot* e do servidor *Mondrian*. Estas ferramentas permitem a implementação de sistemas baseados em OLAP com sofisticadas interfaces sobre a Web, que possibilitam inclusive a geração de gráficos.

Embora as ferramentas *open source* utilizem recursos sofisticados para Web, estas interfaces não são tão amigáveis para utilização pelo usuário final. A configuração das interfaces e a montagem de consultas necessitam o conhecimento de linguagem mdx que é utilizada para manipulação de banco de dados dimensionais. Além disso, as ferramentas não permitem a persistência das consultas após a iteração do usuário.

Neste sentido, foi desenvolvida uma ferramenta *Web open source*, integrada ao servidor OLAP *Mondrian* e a biblioteca *JPivot*, que permite a persistência e criação de consultas sem a necessidade de conhecimento da linguagem mdx. A ferramenta Neuro foi desenvolvida para auxiliar a tomada de decisão através da análise de dados, onde o usuário pode criar as suas consultas e disponibilizá-las à comunidade de interesse.

Este artigo está organizado como segue. A seção 2 apresenta os conceitos e tecnologias utilizadas para o desenvolvimento da ferramenta. A seção 3 apresenta a arquitetura, a modelagem realizada para a ferramenta e as funcionalidades através de experimentos realizados em um estudo de caso. A seção 4 apresenta as conclusões e trabalhos futuros.

2. *Business Intelligence* (BI)

Englobando os conceitos de DW e OLAP está a definição de BI, ou Inteligência de Negócios. [Barbieri 2001] define o conceito de BI, de uma forma mais ampla, como a utilização de variadas fontes de informação para se definir estratégias de competitividade

nos negócios da empresa. Frisando que a diferenciação do modelo BI está relacionada com a forma alternativa de tratamento de informações.

Nas próximas seções serão descritos os conceitos de DW e OLAP, assim como as suas estruturas e elementos.

2.1 Data Warehouse (DW)

O DW é descrito por [Inmon 1997] como uma coleção de dados orientada por assuntos, integrada, variante no tempo e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão. Ainda, para que exista o DW é necessária a extração, transformação e carga, contida no conceito de *Extract, Transform and Load* (ETL). Extração significa basicamente ler e entender as fontes de dados e copiar as partes necessárias para a área de transformação de dados, a fim de serem trabalhadas após a carga [Kimbal 2002].

2.2 On-Line Analytical Processing (OLAP)

Para manipular um armazém de dados (DW) necessita-se de conceitos distintos. Comumente, em base de dados tradicionais para operações convencionais, transacionais, utilizam-se tecnologias chamadas de OLTP (*On-Line Transaction Processing*). Entretanto, no caso de DW, utilizam-se tecnologias denominadas OLAP.

2.2.1 Arquitetura OLAP

As ferramentas OLAP são geralmente de uma das três arquiteturas: ROLAP – *Relational OLAP*, MOLAP – *Multidimensional OLAP* e HOLAP – *Hybrid OLAP*.

A arquitetura ROLAP realiza o processamento do DW em uma estrutura física relacional modelada dimensionalmente. Na arquitetura MOLAP é utilizado o armazenamento físico multidimensional. E a arquitetura HOLAP integra as características de ROLAP e MOLAP, onde os dados atômicos são armazenados em um modelo relacional e os agregados e as dimensões são guardadas em uma matriz n-dimensional [Fernandes e Kantorski, 2008].

2.2.2 Elementos OLAP

As tecnologias OLAP manipulam os dados do DW através de uma estrutura multidimensional denominada cubo. Os cubos representam o DW e são organizados de maneira a realizar consultas rápidas e dinâmicas voltadas ao apoio à decisão. Os cubos são definidos através de dimensões e medidas. As dimensões fornecem as informações descritivas, enquanto que as medidas disponibilizam informações quantitativas que se deseja consultar [Kimball 1996].

As dimensões, ou tabela de dimensões, basicamente compreendem aos aspectos a serem analisados que se relacionam com o cubo, por meio de outra estrutura, a tabela de fatos. [Imhoff, Galemme e Geiger 2003] definem que uma tabela de dimensão é um conjunto de tabelas de referência que provêm a base para restringir e agrupar consultas para a informação em um tabela de fato dentro de um modelo relacional. Estas dimensões possuem membros que, por sua vez, possuem hierarquias. Segundo [*The OLAP Council* 2008] um membro de uma dimensão é um nome ou identificador

discreto usado para identificar uma posição e uma descrição de um item de dado dentro de uma dimensão, enquanto uma hierarquia é o resultado de quaisquer membros de uma dimensão podendo ser organizado com base em relacionamentos pai-filho.

É clara a necessidade de OLAP realizar operações mensuráveis em uma base de dados, logo as medidas possuem considerações como é citado por [Imhoff, Gallemmo e Geiger 2003]. Uma medida é um termo da modelagem multidimensional que se refere a valores, geralmente numéricos, que medem algum aspecto do negócio. As medidas residem em tabelas de fatos.

Representados por tabelas (*fact tables*) os fatos correspondem aos conjuntos de dados mensuráveis e referenciáveis no modelo dimensional. Uma tabela de fato é a tabela dentro de um modelo dimensional que contém as medidas e métricas de interesse [Imhoff, Gallemmo e Geiger, 2003].

2.2.3 Modelagem OLAP

Os cubos possuem suas características e elementos, assim como, possuem a modelagem de suas estruturas. Estas correspondem basicamente a forma como os cubos e suas dimensões são organizados. Os dois modelos mais utilizados e citados na literatura da área são: *Star Schema* (esquema estrela) e *Snowflake Schema* (esquema flocos de neve). O esquema estrela (*star schema*), descrito por [Kimball 2002], é uma representação genérica de um modelo dimensional em um banco de dados relacional em que uma tabela de fatos com uma chave composta é unida a várias tabelas de dimensão, cada uma com uma chave primária. Já o *snowflake schema* (esquema flocos de neve) possui a diferença que as dimensões e as respectivas tabelas de dimensões são normalizadas. Estas características garantem maior organização nos dados manipulados, porém o desempenho na recuperação desses dados será inferior ao modelo *star schema*.

2.2.4 Operações OLAP

As tecnologias e conceitos OLAP caracterizam os cubos de forma detalhada para que sejam manipulados através de consultas e diversas operações, permitindo a visualização destas consultas ao interessado. As principais operações OLAP são *Drill Down e Roll Up, Slice and Dice, Drill across, Ranking e Pivoting*.

2.2 Tecnologias BI

As ferramentas para um ambiente de BI podem ser classificadas como de construção, gerência, uso e armazenamento [Barbieri, 2001], sendo o objeto em foco a informação. Estas classificações são percebidas em conceitos dos itens anteriores, isto é, *Data Warehouse (DW)* e *On-Line Analytical Processing (OLAP)*. Logo, para que um ambiente que faça uso de BI possa beneficiar-se de suas vantagens, deverá utilizar os recursos tecnológicos que apliquem apropriadamente tais conceitos de DW e OLAP.

O Mondrian é um servidor OLAP desenvolvido na linguagem Java sendo parte de um conjunto de tecnologias *open sources*, pertencentes ao projeto Pentaho *Business Intelligence* [Pentaho, 2008]. O JPivot é uma ferramenta que através de suas *tag libraris* permite apresentação e interação com os dados analisados [JPivot, 2008], isto é, visualização e manipulação de tabelas, resultado de consultas OLAP.

Para um melhor entendimento do funcionamento do servidor Mondrian é possível dividi-lo em 4 camadas [Mondrian 2008]: a camada de apresentação (*presentation layer*), a camada dimensional (*dimensional layer*), camada estrela (*star layer*) e a camada de armazenamento (*storage layer*). Conforme [Fernandes e Kantorski 2008], as camadas são definidas como:

- Camada de apresentação: o JPivot é responsável por esta camada, projetada para trabalhar com vários servidores OLAP, especialmente com o Mondrian.
- Camada dimensional: é a camada que recebe, analisa e envia as consultas MDX (*Multidimensional Expressions*).
- Camada estrela: é responsável pela ligação da consulta MDX recebida pela camada anterior e a execução da mesma sobre a base de dados mapeada.
- Camada de armazenamento: executa as consultas solicitadas pela camada estrela sobre a base de dados relacional, logo caracterizando o Mondrian como um servidor ROLAP.

O Mondrian e o JPivot são ferramentas OLAP *open source* e que possuem interface para a web, englobadas pela ferramenta Neuro.

3. A Ferramenta Neuro BI

A ferramenta Neuro BI é parte de um projeto desenvolvido na ULBRA, Campus Santa Maria e na UFSM, composto pelas tecnologias citadas nas seções anteriores que agrega novas funcionalidades para interação com o usuário final. A ferramenta foi desenvolvida na linguagem de programação Java para a Web. Ela executa em um servidor Web e utiliza uma base de dados PostgreSQL para o seu repositório de informações. Além das tecnologias citadas nas seções anteriores, três *frameworks* são utilizados no desenvolvimento da ferramenta: *Sitemesh* (*layout* das páginas), *Hibernate* (manipulação de banco de dados) e *Spring* (lógica de negócios), escolhidos pela disponibilidade de recursos e por serem desenvolvidos em linguagem Java. A figura 1 mostra a arquitetura definida para a ferramenta e a integração com as demais tecnologias. Nas próximas seções serão descritas as características da ferramenta desenvolvida e seus recursos.

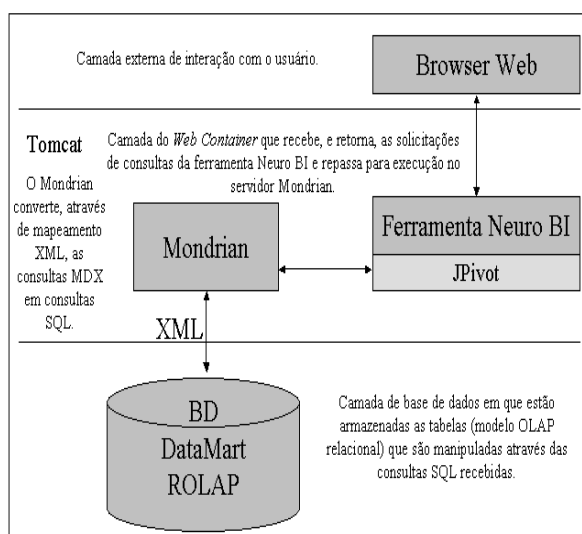


Figura 1. Arquitetura da ferramenta.

3.1 Desenvolvimento

A ferramenta Neuro BI foi desenvolvida utilizando as tecnologias Mondrian e JPivot para agregar recursos que visam facilitar a utilização de recursos de BI pelo usuário final. Desta forma a ferramenta ajuda a tomada de decisão por meio de tecnologia BI sobre uma base de dados pertinente. A Figura 2 mostra uma visão geral dos casos de uso.

O sistema possui autenticação de usuários, através de uma classificação de usuários ou, atores do sistema. a modelagem acima: *usuário*, *usuário-administrador* e *administrador do sistema*.

A seguir são apresentadas as descrições das ações (casos de uso) dos atores do sistema:

- *Usuário*: visualização cubos, categorias e consultas habilitadas;
- *Usuário-administrador*: realização das mesmas operações do *usuário* e ainda alteração de consultas e manipulação para sua exclusiva visualização;
- *Administrador do sistema*: pode acessar as funcionalidades dos usuários anteriores e também possui funcionalidades administrativas do sistema, tais como, controle de usuários, manipulação (criação, alteração e exclusão) dos cubos, categorias e consultas que poderão ser habilitadas para visualização dos respectivos usuários.

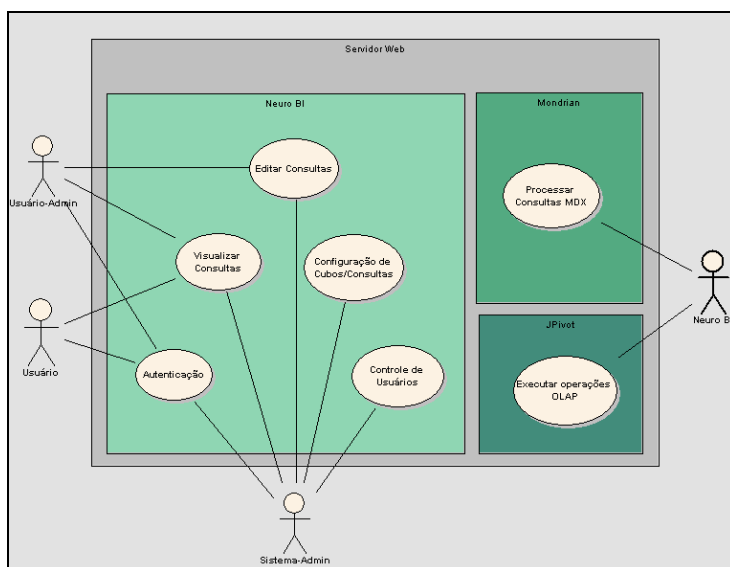


Figura 2. Modelo de Caso de Uso

O Neuro BI possui um repositório para armazenar as informações da aplicação, compreendido pelo modelo de dados apresentado na Figura 3. Estas tabelas estão contidas na base de dados PostgreSQL [PostgreSQL, 2008].

Sendo uma ferramenta OLAP, o Neuro manipula as estruturas OLAP denominadas cubos e consultas MDX (*Multidimensional Expressions*) além de classificar as consultas em várias categorias e habilitar a visualização de consultas para os usuários através de autorizações.

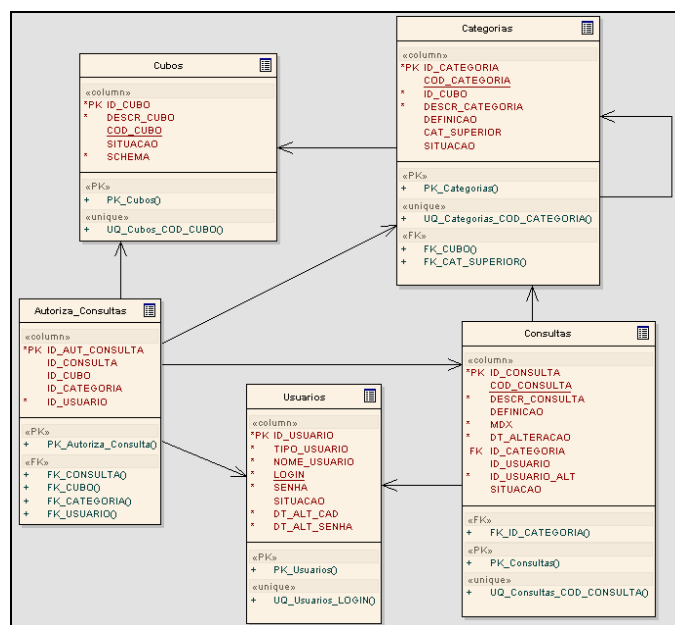


Figura 3. Modelo de Dados Neuro.

3.2 Funcionalidades

O objetivo da ferramenta Neuro é abstrair alguns dos conceitos de tais tecnologias para facilitar sua utilização pelos usuários finais sem a necessidade de conhecimento da linguagem mdx. A seguir, estão descritas as características do Neuro através de interfaces. O estudo de caso realizado para validação é nas informações referentes ao processo seletivo da Universidade Federal de Santa Maria.

Basicamente, quanto as consultas MDX, o Neuro possibilita as operações OLAP comuns (do servidor Mondrian), agregando informações sobre tais consultas, tais como: descrição, título, categoria da consulta e usuário criador (Figura 4). Esta consulta está contida em um cubo, pois este contém as categorias. As categorias possuem, além do título, uma descrição e seus respectivos cubos. Estes possuem a descrição de um item que é necessário para o servidor OLAP, o esquema (*schema*), que é abstraído ao *usuário* e *usuário-administrador*.

Neuro BI | Cubos | Minhas Consultas | Administração | Ajuda | Usuário Administrador | Sair

Cubo Vestibular 2008 - Candidatos, Cursos, Cotas >> Categoria Vestibular 2008 - Candidatos, Cursos, Cotas

Consultas

Listagem de Consultas autorizadas para sua visualização.

Código	Nome	Descrição
C007	Aprovados por Sexo e Sistema de Cotas	Apresenta o total de aprovados por sexo e sistema de cotas. Permite ainda a visualização por opção de curso do candidato aprovado.
C006	Aprovados por Sistema de Cotas	Mostra o total de candidatos aprovados por sistema de cotas. Permite ainda a visualização dos aprovados por curso e sistema de cotas.
C008	Candidatos por Situação e Curso	Mostra o total de candidatos por situação. Total de eusentes, classificados, eliminados. Permite a visualização por opção de curso.
C005	Maior e menor pontuação pos Sistema de Cotas	Mostra a maior e a menor pontuação alcançada pelos candidatos por Sistema de Cotas.
C001	Total de Candidatos Inscritos por Curso	Apresenta o total de candidato inscritos no vestibular UFSM 2008 por curso
C004	Total de Candidatos Inscritos por Sexo e opção de Curso	Mostra o total de candidatos do sexo feminino e masculino por opção de curso no concurso.
C003	Total de Candidatos Inscritos por Sexo e Sistema de Cotas	Mostra o total de candidatos do sexo feminino e masculino por sistemas de cotas.
C002	Total de Inscritos por Sistema de Cotas	Apresenta o total de candidatos inscritos no vestibular 2008 - UFSM por sistema de cotas.

Neuro Business Intelligence v2.0
Centro de Processamento de Dados - UFSM
Sistemas de Informação - Ultra/SM

Atualizado: 25/06/2008 12:23

Figura 4. Listagem de consultas do usuário

O Neuro agrega um importante recurso aos usuários sobre as consultas MDX. Os usuários ao visualizarem uma consulta podem (desde que estejam autorizados) alterar ou criar uma nova consulta, conforme a Figura 5. Deduzindo-se que estes, usuários, desconheçam a linguagem MDX, eles criam e alteram as consultas, apenas manipulando as consultas já existentes (listadas, por exemplo, na Figura 4). É importante salientar que o usuário não necessita alterar o código mdx mostrado na figura 5.

Todos os usuários (*usuário* e *usuário-administrador*), além de previamente identificados, devem possuir autorizações seja para visualizar (Figura 5) os cubos, as categorias ou consultas cadastradas no Neuro. Para isso, há uma seção da ferramenta em que apenas o *usuário-administrador* delega as devidas autorizações.

Neuro BI | Cubos | Minhas Consultas | Administração | Ajuda | Usuário Administrador | Sair

Cadastrar: Consulta

Dados da Consulta

Nome * Total de Inscritos por Sistema de Cotas Código C002

Definição Apresenta o total de candidatos inscritos no vestibular 2008 - UFSM por sistema de cotas. Ativo

MDX *
`select NON EMPTY {[Measures].[inscritos]} ON COLUMNS, NON EMPTY
Hierarchize(Union({[cotas.cotas].[All cotas], [cursos.cursos].[All cursos]}),
Crossjoin([cotas.cotas].[All cotas].Children, {[cursos.cursos].[All cursos]}))
ON ROWS from [Vestibular]`

Categoria * Vestibular 2008 - Candidatos, Cursos, Cotas (*) Dado obrigatório.

Salvar Visualizar Configurações

Neuro Business Intelligence v2.0
Centro de Processamento de Dados - UFSM
Sistema de Informação - UltraSM

Atualizado:
25/06/2008 12:25

Figura 5. Formulário de Consulta.

A Figura 6 mostra a tela de visualização de uma consulta (por um *usuário-administrador*). Na parte superior da figura 6 são mostrados o cabeçalho, a tabela com os dados da consulta e os botões com as operações OLAP (definidas pelo servidor Mondrian). Ainda, entre este cabeçalho e a tabela, estão as informações cadastradas sobre esta consulta visualizáveis opcionalmente através de um botão. À esquerda da tabela está o menu do Neuro, onde pode-se acessar as demais seções da aplicação.

4. Considerações Finais

Este artigo apresentou uma ferramenta OLAP *open source* para apoio a tomada de decisões pelos usuários. A principal contribuição da ferramenta desenvolvida é a possibilidade do usuário criar novas consultas a partir da visualização de todas as dimensões, da tabela de fatos e poder executar as operações OLAP sem a necessidade de conhecer a linguagem mdx. As novas consultas criadas podem ser persistidas em um repositório e serem disponibilizadas para usuários autorizados. A ferramenta desenvolvida permite que usuários do nível gerencial das organizações possam manipular os cubos de dados sem a necessidade de intervenção técnica. A ferramenta pode ser acessada através do *site* <http://portal.ufsm.br:8081/neuro>.

Trabalhos futuros envolvem a integração da ferramenta Neuro com servidores de mapas para facilitar o entendimento dos usuários quando consultas realizadas em bases de dados de cidades forem realizadas.

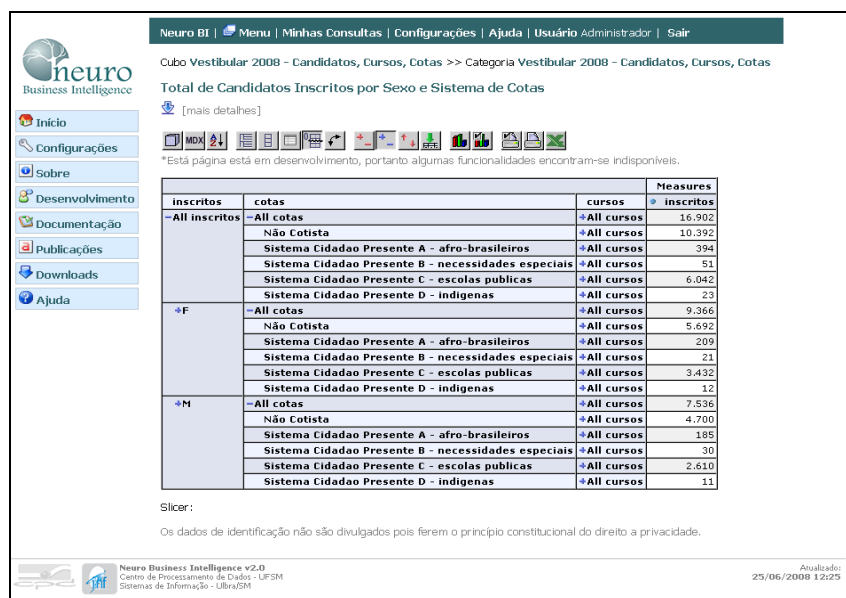


Figura 6. Visualização de uma consulta

Referências

- Ballard, Chuck et al. (2006) “*Dimensional Modeling: In a Business Intelligence Environment. International Business Machines Corporation (IBM Corp.)*”.
<http://www.redbooks.ibm.com/abstracts/SG247138.html>, Junho.
- Barbieri, Carlos. (2001), “*BI – Business Intelligence – Modelagem e Tecnologia*”. Rio de Janeiro: Axcel Books do Brasil.
- Cramer, Renato. (2006), “*Estudo Analítico de Ferramentas Open Source para Ambientes OLAP*”. Criciúma: Universidade do Extremo Sul Catarinense.
- Fernandes, Mikael S.; Kantorski, Gustavo Z. (2008) “*Uma Ferramenta WEB para apoio à Decisão em Ambiente Hospitalar*”. 3ª Conferencia Ibérica de Sistemas y Tecnologías de la Información. Espanha, Universidad de Vigo.
- Imhoff, Claudia; Galemno, Nicholas; Geiger, J. G. (2003), “*Mastering Data Warehouse Design: Relational and Dimensional Techniques*”. Indianapolis: Wiley Publishing, Inc.
- Inmon, W. H. (1997), “*Como construir o Data Warehouse*”. Rio de Janeiro: Campus.
- JPivot. (2008) “*A JSP based OLAP*”, <http://jpivot.sourceforge.net>, Junho.
- Kimball, Ralph; Ross, Margy. (2002), “*The Data Warehouse Toolkit: guia completo para modelagem dimensional*”. Rio de Janeiro: Campus.
- Kimball, Ralph. (1996) “*The Data Warehouse Tollkit*”. John Wiley.
- Mondrian. (2008) “*Mondrian OLAP Server*”, <http://mondrian.sourceforge.net>, Junho.
- Pentaho. (2008) “*Open Source Business Intelligence*”, <http://www.pentaho.org>, Junho.
- PostgreSQL. (2008) “*PostgreSQL*”, <http://www.postgresql.org>, Junho.
- The OLAP Council. (2008) “*The OLAP Council*” <http://www.olapcouncil.org>, Junho.