SAMUEL ZANFERDINI OLIVA

AMBIENTE DE DATA WAREHOUSING PARA INTEGRAÇÃO DE DADOS DE SAÚDE PÚBLICA EM ÂMBITO DE GESTÃO REGIONAL

VERSÃO CORRIGIDA

Ribeirão Preto
2015

SAMUEL ZANFERDINI OLIVA

AMBIENTE DE DATA WAREHOUSING PARA INTEGRAÇÃO DE DADOS DE SAÚDE PÚBLICA EM ÂMBITO DE GESTÃO REGIONAL

VERSÃO CORRIGIDA

Dissertação apresentada ao Programa de Pós–Graduação Interunidades Bioengenharia - Escola de Engenharia de São Carlos / Faculdade de Medicina de Ribeirão Preto / Instituto de Química de São Carlos da Universidade de São Paulo como parte dos requisitos para a obtenção do título de mestre em Ciências.

Área de Concentração: Bioengenharia

Orientador: Prof. Dr. Joaquim Cezar Felipe

Ribeirão Preto
2015

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Oliva, Samuel Zanferdini
O48a AMBIENTE DE DATA WAREHOUSING PARA INTEGRAÇÃO DE DADOS DE SAÚDE PÚBLICA EM ÂMBITO DE GESTÃO REGIONAL / Samuel Zanferdini Oliva; orientador Joaquim Cezar Felipe. São Carlos, 2015.

Dissertação (Mestrado) - Programa de Pós-Graduação Interunidades Bioengenharia e Área de Concentração em Bioengenharia -- Escola de Engenharia de São Carlos; Faculdade de Medicina de Ribeirão Preto; Instituto de Química de São Carlos, da Universidade de São Paulo, 2015.

1. Banco de Dados. 2. Data Warehouse. 3. Sistemas de Informação em Saúde. I. Título.

Programa de Pós-Graduação Interunidades em Bioengenharia
EESC / FMRP / IQSC

Samuel Zanferdini Oliva
Título: "Ambiente de Data Warehousing para integração de dados de saúde pública em âmbito de gestão regional".

DISSERTAÇÃO APRESENTADA AO PROGRAMA DE PÓS - GRADUAÇÃO INTERUNIDADES BIOENGENHARIA - EESC/FMRP/IQSC DA UNIVERSIDADE DE SÃO PAULO PARA OBTENÇÃO DO TÍTULO DE MESTRE EM CIÊNCIAS NA ÁREA DE BIOENGENHARIA.

Aprovado em: 08 / 06 / 2015

Prof. Dr. Joaquim Cezar Felipe (Orientador) — Faculdade de Filosofia Ciências e Letras de Ribeirão Preto - USP

Resultado: Aprovado — Assinatura:

Prof. Dr. Domingos Alves — Faculdade de Medicina de Ribeirão Preto - USP

Resultado: Aprovado — Assinatura:

Prof. Dr. Milton Roberto Laprega — Faculdade de Medicina de Ribeirão Preto - USP

Resultado: Aprovado — Assinatura:

Homologado em: ____/____/____

AGRADECIMENTOS

À minha família por todo o apoio que me proporcionou para que eu conseguisse concluir este projeto.

À minha namorada Mariana por compartilhar comigo os momentos de estudos e pela paciência nos momentos difíceis e em ouvir meus desabafos e preocupações.

Ao meu orientador, Prof. Joaquim Cezar Felipe, por me apoiar e orientar no decorrer do desenvolvimento deste projeto.

Ao Prof. Domingo Alves e ao pessoal do ORAH, que colaboraram com as informações em relação ao Mapa de Saúde e na mediação com o Departamento Regional de Saúde de Ribeirão Preto.

Ao pessoal do Departamento Regional de Saúde de Ribeirão Preto que me permitiu participar das reuniões a respeito do Mapa de Saúde.

Aos gestores que participaram da pesquisa de validação do ambiente desenvolvido: Domingos Alves, Vera Aparecida Felisbino, Janise B. B. Ferreira, Arthur Watanabe, Juan S. Y. Rocha, José Sebastiao dos Santos, Stenio J. C. Miranda, Maria do Carmo. G. G. Caccia-Bava, Milton R. Labrega, Vera Lucia de Jesus. Obrigado pela disponibilidade e pelas valiosas sugestões.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos oferecida para este projeto.

RESUMO

OLIVA, S. Z. Ambiente de *data warehousing* para integração de dados de saúde pública em âmbito de gestão regional. 2015. 87 f. Dissertação (Mestrado). Programa de Pós-Graduação Interunidades Bioengenharia – EESC/FMRP/IQSC, Universidade de São Paulo, Ribeirão Preto.

Estudos relacionados ao aperfeiçoamento na gestão de saúde descrevem casos de sucesso com soluções utilizando a implementação de *Data Warehouse* (DW) como um recurso para facilitar e melhorar a qualidade dos dados para gestão na área de saúde pública. O Mapa de Saúde consiste na descrição da distribuição geográfica de recursos humanos, de ações e de serviços de saúde ofertados pelo SUS e pela iniciativa privada, considerando-se o desempenho aferido a partir dos indicadores de saúde do sistema. A gestão em saúde pública carece de recursos que auxiliem efetivamente os processos de análise e tomada de decisão. O objetivo deste trabalho é contribuir para sanar esta carência, a partir do desenvolvimento de um ferramental de apoio à decisão fundamentado em bases de dados de indicadores de saúde determinados pelo Mapa de Saúde. Este ferramental consiste no planejamento e na implementação de um ambiente de *Data Warehousing* para integração de bases de dados legadas, oriundas dos documentos gerados pelas Redes Regionais de Atenção à Saúde (RRAS) e coletadas pela Secretaria de Saúde do Estado de São Paulo, fornecendo aos usuários gestores os recursos necessários para a geração de consultas ad-hoc complexas com flexibilidade e rapidez sobre os dados contemplados pelo Mapa de Saúde. As fases do processo de desenvolvimento são sete, a saber: (1) Análise dos dados definidos pelo Mapa de Saúde e disponibilizados pelas fontes públicas; (2) Modelagem e implementação de um banco de dados integrador; (3) Modelagem dimensional e implementação do DW; (4) Criação do processo de extração, transformação e carga; (5) Geração dos metadados e configuração do servidor de análises; (6) Análises; (7) Testes. Como resultado, todas as fases de desenvolvimento do ambiente de *Data Warehousing* foram cumpridas, bem como o procedimento de avaliação junto aos gestores. De acordo com a avaliação, o sistema cumpre de forma satisfatória o seu objetivo de tornar o processo de gestão de saúde mais efetivo, flexível e produtivo.

Palavras chaves: banco de dados; *data warehouse*; sistemas de informação em saúde.

ABSTRACT

OLIVA, S. Z. Data Warehousing environment for public healthcare data integration on the regional management scope. 2015. 87 f. Program of Postgraduate Interunits Bioengineering – EESC/FMRP/IQSC. University of São Paulo. Ribeirão Preto.

Studies related to the enhancement on health management describe instances of success with solutions concerning the Data Warehouse (DW) implementation as a resource to facilitate and improve data quality for management on health public area. The Health Map consists of the geographical description of human resources, actions and services distribution of health offered by the Brazilian National Health System and by the private sector, taking into account the existent installed capacity, the investment, and the performance assessed from system's health indicators. The goal of this project is to develop a decision support toolset based on health indicators databases defined by the Health Map. This toolset consists of the planning and the Data Warehousing environment implementation for the integration of legacy databases, arising from documents generated by the Regional Health Care Networks and collected by the São Paulo State Health Department, hence providing for users a necessary suite of tools for complex ad-hoc queries generation with flexibility and celerity over the data contemplated by the Health Map. Development process steps are seven, namely: (1) Analysis of data defined by the Health Map and available by public sources; (2) Integrator Database modeling and implementation; (3) DW dimensional modeling; (4) Creation of the extraction, transformation and load process; (5) Metadata generation and analysis server configuration; (6) Analysis; (7) Tests. As a result, the Data Warehousing environment development process steps were completely fulfilled, as well as tests and evaluation steps along with the health managers.

Keywords: database; data warehouse; healthcare information systems

SUMÁRIO

1. Introdução ____ *12*

2. Motivação e Objetivos ____ *17*

3. Base Conceitual ____ *18*

3.1. Data Warehousing (DWing) ____ **19**

3.1.1. Arquitetura ____ 20

3.1.2. Organização dos Dados ____ 22

3.2. ODS ____ **23**

3.3. Metadados ____ **23**

3.4. Esquema Estrela ____ **24**

3.5. OLAP ____ **26**

4. Trabalhos Correlatos ____ *27*

4.1. Implementações de DWing ____ **27**

4.2. Ferramentas de Desenvolvimento de DWing ____ **31**

5. Desenvolvimento do Projeto ____ *34*

5.1. Proposta de Arcabouço Computacional ____ **34**

5.2. Etapas do Processo de Desenvolvimento ____ **36**

5.3. Ferramentas de Desenvolvimento ____ **38**

6. Resultados ____ *40*

6.1. Análise dos dados definidos pelo Mapa de Saúde ____ **40**

6.2. Modelagem do Banco de Dados Integrador ____ **40**

6.3. Modelagem Dimensional do DW ____ **43**

6.3.1. Esquema Estrela de Mortalidade ____ 44

6.3.2. Esquema Estrela de Morbidade ____ 45

6.3.3. Esquema Estrela de Mortalidade por Causas Específicas ____ 46

6.3.4. Esquema Estrela de Amostra de Água ____ 47

6.3.5. Esquema Estrela de Equipamento Hospitalar ____ 47

6.3.6. Esquema Estrela de Leito Hospitalar 48
6.3.7. Esquema Estrela de Análise Profissional 49
6.3.8. Constelação de Fatos 50

6.4. Criação do Processo ETL 54
6.4.1. Carga do Esquema Estrela de Mortalidade 55
6.4.2. Carga do Esquema Estrela de Equipamento Hospitalar 59

6.5. Configuração do Servidor OLAP 61

6.6. Exemplos de Análises 64

6.7. Validação 69

7. Discussão 71

7.1. Aquisição de fontes legadas 72

7.2. Avaliação de Ferramentas 73

7.3. Validação 73

8. Conclusão 74

9. Trabalhos Futuros 75

Referências Bibliográficas 76

APÊNDICE A - Operações OLAP Utilizando o JPivot do Servidor PBA 80

Lista de Figuras

Figura 1 - Redes de Atenção à Saúde e respectivas DRS e Regiões de Saúde do Estado de São Paulo (SECRETARIA DE ESTADO DA SAÚDE, 2014). 14
Figura 2 - RRAS 13 e respectivos DRS, Regiões de Saúde e Municípios (SECRETARIA DE ESTADO DA SAÚDE, 2014). 15
Figura 3 – Página do ORAH exibindo tabelas contidas no Mapa de Saúde. 16
Figura 4 – Arquitetura de DW traduzida e adaptada de Kimball e Ross, 2002. 20
Figura 5 - Estrutura de um DW (CIFERRI, 2002). 23
Figura 6 - Esquema estrela e suas principais características. 25
Figura 7 - Componentes do arcabouço computacional GDMS. 35
Figura 9 – Diagrama de Componentes do *Pentaho BI Suite* traduzido e adaptado de BOUMAN e DONGEN (BOUMAN; DONGEN, 2009). 39
Figura 10 – DER parcial do BD integrador do GDMS. 42
Figura 11 - DER do Esquema Estrela de Mortalidade. 44
Figura 12 - DER do Esquema Estrela de Morbidade. 45
Figura 13 - DER do Esquema Estrela de Mortalidade por Causas Específicas. 46
Figura 14 - DER do Esquema Estrela de Amostra de Água. 47
Figura 15 - DER do Esquema Estrela de Equipamento Hospitalar. 48
Figura 16 - DER do Esquema Estrela de Leito Hospitalar. 49
Figura 17 - DER do Esquema Estrela de Análise Profissional. 50
Figura 18 – DER da Constelação de Fatos por Municípios. 52
Figura 19 – DER da Constelação de Fatos por Hospital. 53
Figura 20 - Representação gráfica de uma transformação no PDI - *Spoon*. 54
Figura 21 - Representação gráfica de uma tarefa no PDI - *Spoon*. 55
Figura 22 - Processo de carga dos dados da dimensão “Dim_Municipio”. 55
Figura 23 - Processo de geração dos dados da dimensão "Dim_Ano". 56
Figura 24 - Processo de carga dos dados da dimensão "Dim_CID10". 57
Figura 25 - Processo de carga dos dados da dimensão "Dim_Faixa_Etaria". 57

Figura 26 - Processo de carga dos dados da dimensão "Dim_Sexo". 58
Figura 27 - Processo de carga dos dados da tabela de fatos "Fato_Mortalidade". 58
Figura 28 - Processo de carga dos dados da dimensão "Dim_Hospital". 59
Figura 29 - Processo de carga dos dados da dimensão "Dim_SUS". 60
Figura 30 - Processo de carga dos dados da dimensão "Dim_Equipamento". 60
Figura 31 - Processo de carga dos dados da tabela de fatos "Fato_Equipamento". 61
Figura 32 - Interface gráfica do *Pentaho Schema Workbench*. 62
Figura 33 - PSW com o item do cubo de dados “Profissional” expandido. 63
Figura 34 - Cubo "Profissional" em XML. 64
Figura 35 – Análise construída do cubo "Profissional". 65
Figura 36 – Início da “Análise de Mortalidade por Capítulos CID10”. 66
Figura 37 – Aplicação de filtros na “Análise de Mortalidade por Capítulos CID10”. 67
Figura 38 – Resultado da “Análise de Mortalidade por Capítulos CID10” e seu respectivo gráfico. 68
Figura 39 – Análise construída da constelação de fatos por município. 69
Figura 40 - Gráfico de barras das pontuações das médias para as questões. 71
Figura 41 - Interface inicial do aplicativo Web do servidor PBA. 80
Figura 42 – Análise do cubo “Mortalidade_Natureza”. 81
Figura 43 – Exemplo da operação *drill-member*. 82
Figura 44 – Exemplo da operação *drill-position*. 83
Figura 45 – Exemplo de consulta com a operação *drill-replace*. 84
Figura 46 – Exemplo de aplicação de filtro na análise do cubo “Mortalidade_Natureza”. 85
Figura 47 – Exemplo da operação *swap axes* e de criação de gráfico. 86
Figura 48 – Menu de opções do JPivot. 86

Lista de Tabelas

Tabela 1 - Comparativo de plataformas BI *open source* 32
Tabela 2 - Mapeamento de informações dos BDpub-GDMS → BDint-GDMS 43

1. Introdução

A gestão na área da saúde pública, no que diz respeito ao processo de tomada de decisão, busca definir políticas e ações que possibilitem o aperfeiçoamento do atendimento à saúde para a população. Para que isto seja efetuado com sucesso, é importante que os gestores estejam munidos de informações atualizadas e relevantes sobre os diferentes problemas.

Estudos relacionados à melhoria na gestão de saúde descrevem casos de sucesso com soluções relacionadas à implementação de *Data Warehouse* (DW) como um recurso para facilitar e aumentar a qualidade dos dados para gestão na área de saúde pública. O DW é um banco de dados destinado a auxiliar a tomada de decisões em gestão de negócios, que caracteriza-se por ser integrado, baseado em assunto, não volátil e estável em relação ao tempo (INMON, 2005). Como exemplos de aplicações em saúde, podemos citar EINBINDER et al. (2001) que apresentam o estudo e a construção de um repositório de dados que consiste em um DW para suporte a pesquisa e educação em um centro médico acadêmico, com a função também de fornecer dados para gestores e administradores. Outro exemplo encontra-se em (EVANS; LLOYD; PIERCE, 2012), que descreve um estudo de caso sobre um DW empresarial aplicado a registros médicos, permitindo que dados de uma organização com numerosos registros de internações e unidade ambulatoriais sejam integrados e analisados.

Os objetivos de um DW estão relacionados às preocupações na gestão de negócios de qualquer organização em relação à grande quantidade de dados e à dificuldade de acesso a eles. Desta forma, os seguintes interesses são os requisitos para um DW: tornar as informações da organização facilmente acessíveis; apresentar as informações da organização de maneira consistente; ser adaptativo e flexível a mudanças; ser uma base segura que protege a informação; servir como um alicerce para melhoria da tomada de decisão (KIMBALL; ROSS, 2002).

CHAUDHURI e DAYAL (1997) definem o *Data Warehousing* (DWing) como uma coleção de tecnologias de apoio a decisão, destinada em habilitar o trabalhador intelectual (executivo, gestor, ou analista) a tomar decisões melhores e mais rapidamente, baseada na implementação de um DW.

A implantação de um ambiente de DWing exige a integração dos dados oriundos de diferentes fontes de informações legadas, a partir de ambientes transacionais internos e fontes de dados externas, para uma base de dados especializada mais ampla. Sua construção fornece uma visão abrangente de toda a organização a partir do acesso a dados históricos, confiáveis para facilitar a tomada de decisão, com o mínimo possível de sobrecarga nos sistemas transacionais (POE; BROBST; KLAUER; POE, 1997).

A presente proposta consiste na implementação de um ambiente de DWing baseado em informações de saúde pública. Para tal, a referência em termos de conjunto de informações é o chamado Mapa de Saúde. O Mapa de Saúde foi criado pelo Ministério da Saúde, sob o decreto 7.508 de 28 de junho de 2011 (PORTAL DA LEGISLAÇÃO GOVERNO FEDERAL, 2014), e consiste na descrição geográfica da distribuição de recursos humanos e de ações e serviços de saúde ofertados pelo SUS e pela iniciativa privada, considerando-se a capacidade instalada existente, os investimentos e o desempenho aferido a partir dos indicadores de saúde do sistema. Deve ser utilizado na identificação das necessidades de saúde com o objetivo de orientar o planejamento integrado dos entes federativos, contribuindo para o estabelecimento de metas da saúde.

Em suma, o Mapa de Saúde é um documento composto por um conjunto de capítulos, de forma que cada capítulo representa um ou mais indicadores de saúde, a saber:

I. Condições geográficas, demográficas, e sócio-econômicas;
II. Perfil de morbimortalidade;
III. Atenção básica como ordenadora;
IV. Vigilância em saúde;
V. Estrutura (capacidade instalada, equipamentos e assistência);
VI. Redes temáticas;
VII. Regulação;
VIII. Controle social;
IX. Gestão do trabalho e educação em saúde;
X. Financiamento do sistema;
XI. Judicialização;
XII. Transporte sanitário; e

XIII. Recursos humanos.

As informações contidas nestes capítulos estão separadas por seções e/ou subseções. Dentro de cada seção e/ou subseção estão especificadas descrições textuais dos indicadores, bem como tabelas contemplando os valores numéricos relativos aos diversos indicadores. Os documentos do Mapa de Saúde são construídos para um ano específico, algumas informações contidas neles são equiparadas com outros anos antes de seu ano de construção para fins de cálculos e comparações.

As informações contidas nas fontes originais são dadas em níveis de detalhes heterogêneos - municípios, regiões de saúde, Departamentos Regionais de Saúde (DRS) e Estados. A Figura 1 apresenta a divisão geográfica do Mapa da Saúde do estado de São Paulo, com suas respectivas Redes Regionais de Atenção à Saúde (RRAS), Departamentos Regionais de Saúde (DRS) e Regiões de Saúde.

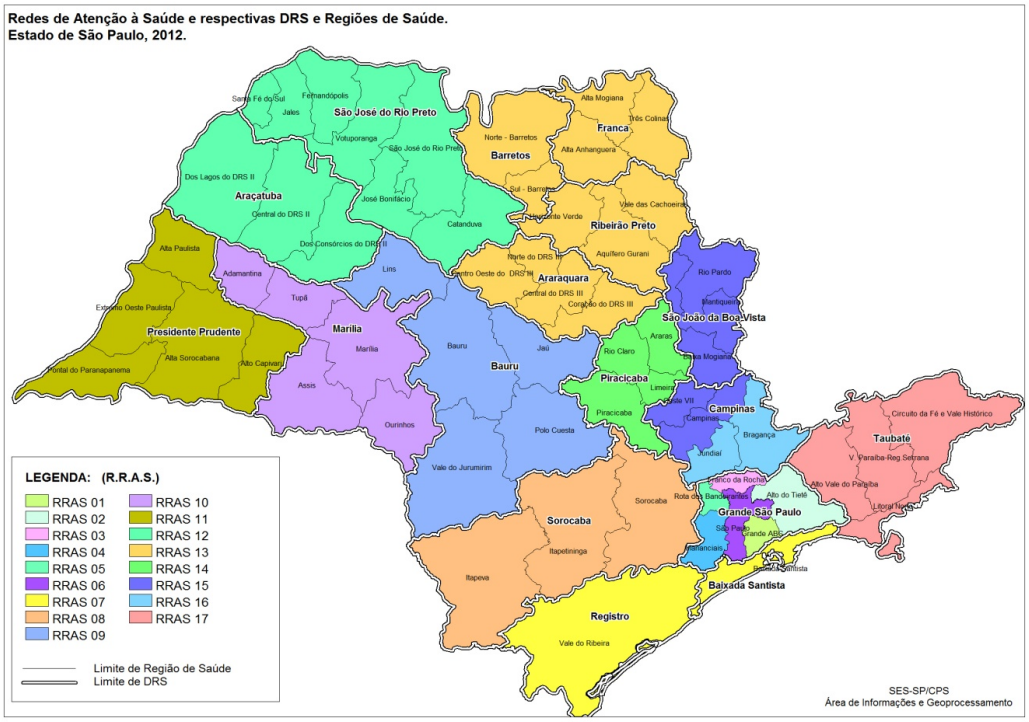


Figura 1 - Redes de Atenção à Saúde e respectivas DRS e Regiões de Saúde do Estado de São Paulo (SECRETARIA DE ESTADO DA SAÚDE, 2014).

A Figura 2 apresenta a RRAS 13 e a divisão das áreas geográficas com seus respectivos DRS, Regiões de Saúde e Municípios.

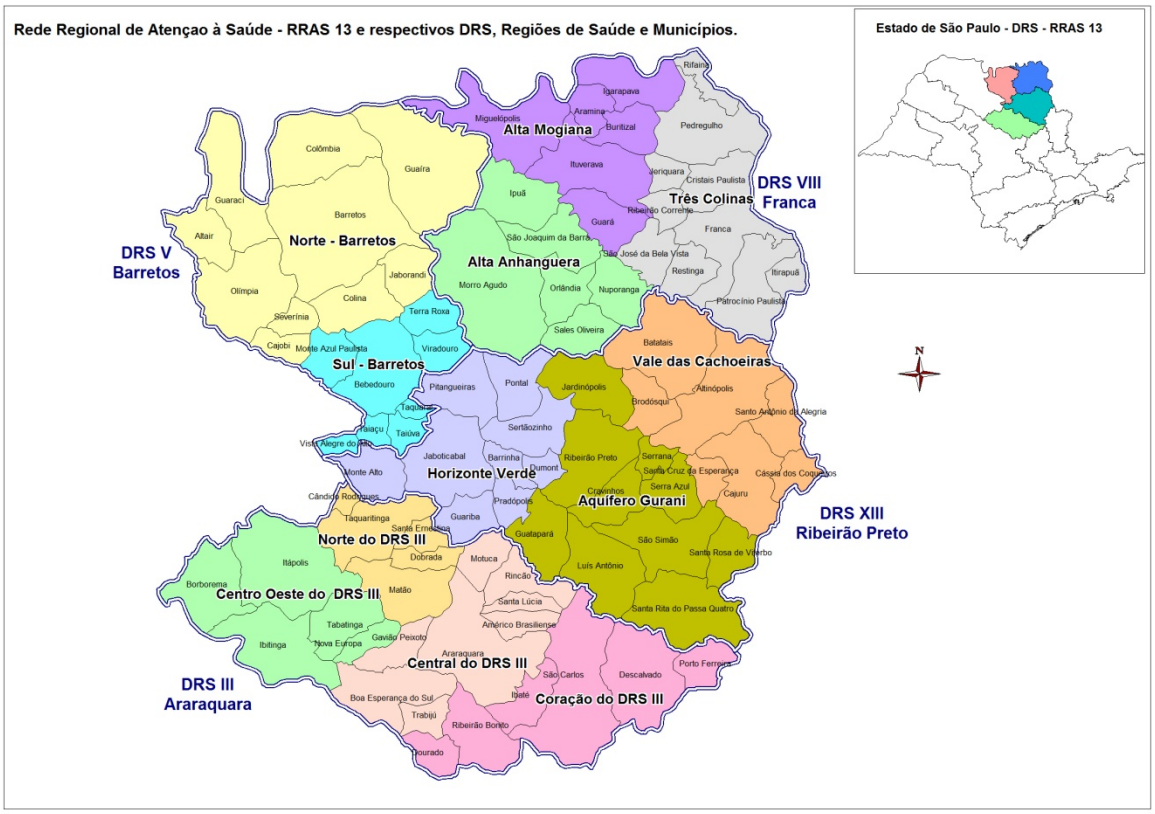


Figura 2 - RRAS 13 e respectivos DRS, Regiões de Saúde e Municípios (SECRETARIA DE ESTADO DA SAÚDE, 2014).

A coleta das informações que comporão o mapa de saúde regional é atribuição dos Departamentos Regionais de Saúde. O Departamento Regional de Saúde da Região de Ribeirão Preto (DRS XIII), do estado de São Paulo, realizou, até o presente momento, uma única coleta, compreendendo o ano de 2010, e agregou estas informações em tabelas, de acordo com a distribuição nos capítulos citados. A fim de apoiar este processo, o Observatório Regional de Assistência Hospitalar (ORAH) (ALVES, 2015) da Faculdade de Medicina de Ribeirão Preto (FMRP) da Universidade de São Paulo desenvolveu uma página Web que permite a navegação entre os capítulos, apresentando as respectivas tabelas estáticas para os diferentes indicadores para o ano de 2010. A fim de exemplificar a estrutura, a Figura 3 mostra o Mapa de Saúde (apresentado pela página Web do ORAH) com o documento referente à região de saúde Aquífero

do Guarani e seus capítulos. O último capítulo, de recursos humanos, foi expandido de modo que possamos visualizar as subseções, que contabilizam seis. A última seção, a seção seis, foi selecionada e, assim, nós podemos visualizar os indicadores de saúde do lado direito da página, nas suas respectivas tabelas.

A equipe do ORAH apoiou também o desenvolvimento do presente trabalho, fornecendo informações e participando das discussões iniciais sobre a estrutura do arcabouço proposto (OLIVA *et al.*; 2014).

ORAH - Observatório Regional de Atenção Hospitalar
Mapa de Saúde

Esconder tudo | Expandir tudo
- RRAS-13
  - Regiões de Saúde
    - Região de Saúde Aquífero Guarani
      - A. Equipe Técnica Responsável
      - B. Introdução
      - I. Condições Geográficas, Demográficas e Sócio-Econômicas
      - II. Perfil de Morbimortalidade – segundo residência
      - III. Atenção Básica como Ordenadora do Sistema
      - IV. Vigilância em Saúde
      - V. Estrutura – capacidade instalada, equipamentos e assistência
      - VI. Redes Temáticas
      - VII. Regulação
      - VIII. Controle Social
      - IX. Gestão do Trabalho e Educação na Saúde
      - X. Financiamento do Sistema
      - XI. Judicialização
      - XII. Transporte Sanitário
      - XIII. Recursos Humanos
        - 1. Ocupações que atendem ao SUS – nível superior
        - 2. Ocupações que atendem ao SUS – nível médio
        - 3. Ocupações alocadas na Atenção Básica (UBS/PSF)
        - 4. Ocupações alocadas na Atenção Especializada ambulatorial que atendem ao SUS
        - 5. Ocupações alocadas na Atenção hospitalar que atendem ao SUS
        - 6. Ocupações que não atendem ao SUS

Ocupações que Não Atendem ao SUS

Tabela 1 - Número e coeficiente por mil habitantes de médicos por especialidade e total – NÃO SUS, na região de saúde Aquífero Guarani.

Categoria profissional	Nº de profissionais da categoria	Coeficiente por mil habitantes
Médico (Anestesista)	63	0,08
Médico (Cirurgião Geral)	50	0,06
Médico (Clínico Geral)	183	0,23
Médico (Gineco Obstetra)	264	0,33
Médico (de Família)	1	0,00
Médico (Pediatra)	232	0,29
Médico (Psiquiatra)	80	0,10
Médico (Radiologista)	61	0,08
Médico (Outras Especialidades)	1	0,00
Total de Médicos	**935**	**1,16**

Fonte: CNES 2012

Tabela 2 - Número e coeficiente por mil habitantes de médicos por especialidade e total – NÃO SUS, do DRS XIII - Ribeirão Preto.

Categoria profissional	Nº de profissionais da categoria	Coeficiente por mil habitantes
Médico (Anestesista)	82	0,06
Médico (Cirurgião Geral)	79	0,06
Médico (Clínico Geral)	347	0,26
Médico (Gineco Obstetra)	358	0,27
Médico (de Família)	1	0,00
Médico (Pediatra)	367	0,28
Médico (Psiquiatra)	106	0,08
Médico (Radiologista)	75	0,06
Médico (Outras Especialidades)	2	0,00
Total de Médicos	**1417**	**1,07**

Fonte: CNES 2012

Figura 3 – Página do ORAH exibindo tabelas contidas no Mapa de Saúde.

2. Motivação e Objetivos

As fontes de informações de saúde, especificamente as informações sobre assistência hospitalar, distribuição de recursos humanos, de ações e serviços de saúde ofertados pelo SUS e pela iniciativa privada, e outras obtidas a partir dos indicadores de saúde do sistema e descritas por meio de divisões de áreas geográficas pelo Mapa de Saúde, estão atualmente organizadas de maneira estática em documentos textos, planilhas, e em bases de dados oficiais sobre informações em saúde.

A gestão da saúde requer o planejamento e a organização estratégica para a coleta, o processamento, o armazenamento, a distribuição e a avaliação das informações inerentes aos recursos, ações e serviços oferecidos da saúde com o propósito de auxiliar e aperfeiçoar o atendimento à população. Neste contexto, a gestão da informação de saúde oferece uma forma de organizar e disponibilizar o conhecimento sobre informações da saúde combinado com aspectos estratégicos e com a tecnologia da informação.

O sistema público de saúde brasileiro apresenta sérias dificuldades na gestão da informação e uma ausência de planejamento para atendimento de suas demandas e de controle de seus recursos considerando um intervalo de tempo restrito (BORGES, 2014). Dentre estas dificuldades podemos citar, por exemplo, a complexidade na gestão de recursos humanos (SCHERER; PIRES; SCHWARTZ, 2009), o desafio na gestão de materiais, de tecnologia, e de bens financeiros (FEUERWERKER; CECÍLIO, 2007). Considerando especificamente a proposta do Mapa de Saúde e a forma tabular estática como atualmente as suas informações são disponibilizadas, os gestores que colaboraram com este trabalho citam uma série de restrições relevantes, tais como: ausência de padronização das informações coletadas de fontes heterogêneas; dificuldade de realizar análises sobre distribuições temporais; impossibilidade de manipular os valores dos indicadores em diferentes níveis de agregação espacial (municipal, regional, estadual); inexistência de ambientes customizados de acordo com as necessidades de cada tipo de gestor.

A proposta de ambiente DWing tem como motivação melhorar a disponibilidade e a acessibilidade dessas informações descritas pelo Mapa de Saúde, as quais estão atualmente

dispersas em diversos documentos e sistemas legados de informação da saúde e, dessa forma, tornar essas informações dispostas de forma estáticas acessíveis com flexibilidade, praticidade e rapidez. Esse ambiente visa capacitar os gestores em níveis estadual, regional e municipal a construírem consultas ad-hoc complexas de uma maneira dinâmica por meio de um ferramental que possibilita obter uma ampla variedade de visões em relação à agregação e ao detalhamento dos dados de saúde.

Portanto, as informações descritas no Mapa de Saúde alinhadas com o conceito de DWing fornecerão um sistema de apoio à tomada de decisão para monitoramento e planejamento na gestão em saúde.

A fim de viabilizar de forma otimizada os objetivos do Mapa de Saúde no que diz respeito ao monitoramento e à análise da informação, este projeto de mestrado consiste em desenvolver um arcabouço de apoio à decisão fundamentado em bases de dados de indicadores de saúde, que consiste no planejamento e na implementação de um ambiente de DWing para integração de bases de dados legadas, oriundas dos documentos gerados pelas Redes Regionais de Atenção à Saúde (RRAS) e coletadas pela Secretaria de Saúde do Estado de São Paulo, fornecendo aos usuários o ferramental necessário para a geração de consultas ad-hoc complexas com flexibilidade e rapidez sobre os dados contemplados pelo Mapa de Saúde.

Este desenvolvimento consiste basicamente de sete atividades principais, a saber: análise dos dados definidos pelo Mapa de Saúde; modelagem e implementação de um banco de dados integrador; modelagem dimensional do DW; criação do processo ETL; geração de metadados e configuração do servidor; análises; e testes. Todo o projeto estará respaldado no uso de padrões de informação em saúde previstos pelo Sistema Único de Saúde brasileiro.

3. Base Conceitual

Esta seção aborda os principais conceitos envolvidos na implementação e desenvolvimento de um DW.

3.1. Data Warehousing (DWing)

Conforme Kimball e Caserta (2004, p. 23), um “DW é um sistema que extrai, limpa, formata, e entrega dados fontes em um repositório de dados dimensionais e, assim, apoia e implementa consultas e análises para o propósito de tomada de decisões". O conjunto de dados contidos em um DW caracteriza-se como orientado a assunto, integrado, não volátil e histórico com a função de apoiar em decisões de gerenciamento (INMON, 2005).

“Orientado a assunto” significa que o DW é focado nos temas de negócio de maior interesse da corporação. Como, por exemplo: produtos, clientes, contas, política. “Integrado” indica que os dados são obtidos do ambiente transacional, geralmente de múltiplas fontes distintas e então são integrados em uma base única, o DW. Enquanto os dados são extraídos, os mesmos são reformatados para eliminar possíveis inconsistências, resequenciados e sumarizados. “Não volátil” indica que os dados do DW são carregados com uma determinada frequência e acessados, mas não são sobrepostos nas operações de atualização. Ou seja, os dados permanecem estáveis por longo período de tempo. Quando ocorrem mudanças nas bases de dados fontes, um novo carregamento dos dados deve ser realizado não sobrepondo os dados anteriores. Dessa forma, é mantido o histórico dos dados no DW. “Histórico” implica que os dados são relevantes a um grande espectro de tempo. Os dados possuem marcações de tempo, sendo que para cada nova entrada de registro no DW é associada, de forma implícita ou explícita, um componente de tempo. Assim é possível realizar análises temporais sobre as informações e encontrar tendências através de dados históricos.

O DWing é um processo de extração de dados de fontes legadas, bem como de banco de dados transacionais. Neste, as informações são integradas e transformadas para uma estrutura organizada em um formato amigável, que visa encorajar a análise de dados e fornecer suporte a tomada de decisões baseadas em fatos. O processo de DWing envolve uma coleção de ferramentas, repositórios de dados e metadados que em conjunto formam um ambiente de funcionamento para a utilização do DW.

3.1.1. Arquitetura

O DWing é constituído pela combinação de componentes os quais o tornam um ambiente de armazenamento completo. Cada componente possui uma regra e uma função específica. A Figura 4 apresenta, segundo Kimball e Ross (KIMBALL; ROSS, 2002)**,** os quatro componentes elementares para o funcionamento do ambiente de DWing, os quais são: fontes de dados operacionais, área de preparação dos dados, área de apresentação dos dados e ferramentas de acesso a dados.

O primeiro componente consiste nos **sistemas operacionais fontes**, os quais contêm diversos dados heterogêneos com estruturas e formatos relacionais e não relacionais, além de arquivos e bancos de dados utilizando estruturas distintas e complexas (BONTEMPO; ZAGELOW, 1998).

O próximo componente é a **área de preparação dos dados**, onde ocorre tanto o armazenamento (arquivos ou tabelas relacionais) como também um conjunto de processos, conhecido como **ETL (*extraction-transformation-load*)**, ou seja, extração, transformação e carga.

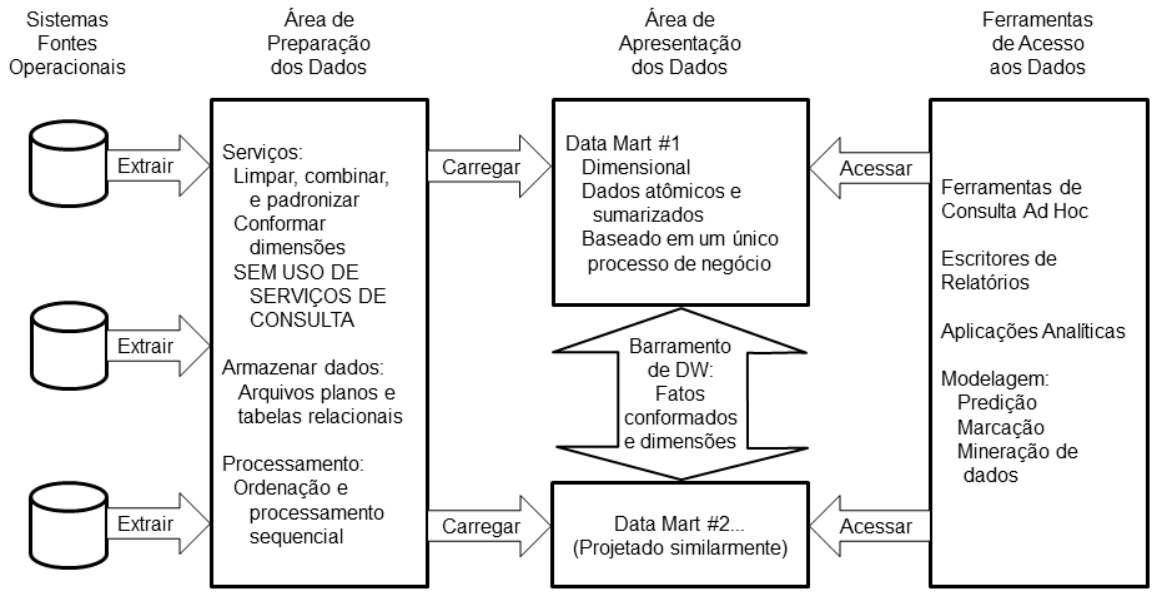


Figura 4 – Arquitetura de DW traduzida e adaptada de Kimball e Ross, 2002.

A **extração** é o primeiro passo do conjunto de processos, o qual consiste na leitura e no entendimento das fontes de dados e a cópia desses dados para a área de preparação dos dados. Uma vez que os dados estão na área de preparação dos dados, ocorrem as transformações.

A **transformação** implica nas correções de inconsistências dos dados, tais como limpeza dos dados (corrigindo conflitos de domínio, lidando conflitos de nome, ou analisando padrões de formatos), combinação de dados de múltiplas fontes, eliminação de redundâncias de dados e atribuição de chaves para o armazenamento.

Por fim, no terceiro passo, ocorre a **carga,** na qual os dados devem ser entregues para a área de apresentação dos dados. Neste passo também podem ocorrer pré-processamentos adicionais, tais como: verificação de restrições de integridade, ordenação, sumarização e agregação, construção de índices e particionamento para múltiplas áreas de armazenamento (KIMBALL; ROSS, 2002).

Após o processo ETL, as informações são transferidas para a **área de apresentação dos dados**. Nessa área é realizada a organização, o armazenamento e a disponibilização de consultas, disponibilização de relatórios, e outras aplicações analíticas. Geralmente a área de apresentação dos dados é referida como uma série de ***data marts*** integrados, sendo que os *data marts* representam subconjuntos de dados do DW classificados segundo os temas selecionados de um determinado negócio (CHAUDHURI; DAYAL, 1997). Os *data marts* devem conter dados detalhados e atômicos, enquanto também podem conter dados sumarizados e agregados. Com o conceito de data marts, o DW pode ser compreendido como sendo a base de dados que representa a integração dos mesmos. Diversos *data marts* são construídos de maneira isolada para depois serem integrados e formar um amplo DW (SAMOS *et al.*, 1998).

O ultimo componente são as **ferramentas de acesso a dados**, que contêm as ferramentas de consulta *ad hoc*, geradores de relatórios, cubos OLAP, além de aplicações analíticas, tais como técnicas de *data mining* (mineração de dados); *forecasting* (previsão) e *scoring* (pontuação) (KIMBALL; CASERTA, 2004).

3.1.2. Organização dos Dados

Os dados em um ambiente DW estão organizados em diferentes níveis de detalhes. No nível de detalhes mais baixo estão os dados obtidos do ambiente operacional. Devido ao crescimento da quantidade de dados coletados periodicamente, ocorre o processo de transferência dos dados contidos no nível inferior de volta para o nível que contém as bases de originais, mantendo a granularidade dos dados durante esse processo, chamado de envelhecimento. Na medida em que os dados são sumarizados eles são passados do nível inferior para os níveis intermediários, que representam agregações crescentes dos dados armazenados. Por fim, há o nível superior no qual os dados são altamente sumarizados (CIFERRI, 2002). A Figura 5 apresenta a organização dos dados de acordo com o nível de detalhamento das informações presente no DW.

A granularidade diz respeito ao nível de detalhe ou agregação do armazenamento no DW. Quanto maior o nível de detalhe, menor é a granularidade. Esta é uma das questões mais críticas em um projeto de DW porque afeta o volume dos dados e o tipo de consultas que podem ser respondidas pelo DW. Quando um nível de granularidade muito alto é utilizado, o volume do DW é reduzido, juntamente com o número de consultas que podem ser respondidas (INMON, 2005).

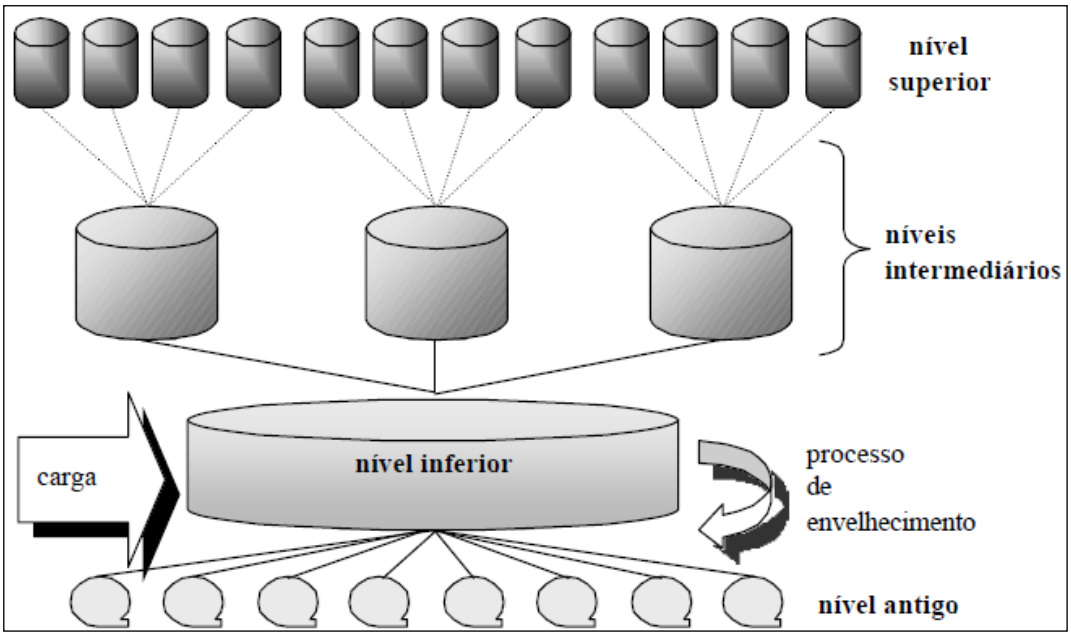


Figura 5 - Estrutura de um DW (CIFERRI, 2002).

3.2. ODS

O ODS (*operational data store*), ou armazém de dados operacionais, é um banco de dados para sistemas de processamento transacionais utilizado para fornecer dados limpos e formatados (GRAY; WATSON, 1998). c

3.3. Metadados

"Os metadados (dados sobre dados) são definidos como qualquer informação necessária para desenvolver, manter e explorar um sistema de DW" (VADUVA; DITTRICH, 2001).

Os metadados podem ser classificados em metadados de gestão dos dados e metadados de acesso aos dados. Os metadados de gestão dos dados relacionam-se com as instruções para as tarefas de extração, limpeza e carga. Esses metadados auxiliam o administrador do DW com o processo ETL e também é de interesse para determinados usuários com informações sobre a origem dos dados. Por outro lado, os metadados de acesso aos dados beneficiam os usuários

finais do sistema de DW funcionando como um dicionário de conteúdo dos dados (KIMBALL; CASERTA, 2004). Além disso, os metadados podem pertencer a três categorias: metadados administrativos, metadados específicos da aplicação e metadados de auditoria.

Os metadados administrativos contêm informações sobre o esquema do DW, que consiste do esquema lógico, o esquema físico, extração dos dados, limpeza e regras de transformação, condições de recarregamento dos dados e o mapeamento lógico entre o esquema lógico e físico. Os metadados administrativos também possuem informações sobre o mapeamento da semântica dos dados, informações sobre perfis de usuários, e do esquema conceitual do DW. Os metadados específicos da aplicação dispõem de um conjunto de terminologias específicas para o domínio da aplicação, restrições de aplicações e outras políticas. Os metadados de auditoria incluem informações relacionadas a relatórios de erros, estatísticas do ambiente de *DWing* e auditoria de experimentações (WU; BUCHMANN, 1997).

3.4. Esquema Estrela

A estrutura de um DW é geralmente representada utilizando um esquema estrela ou floco de neve, baseado em uma visão multidimensional dos dados, consistindo de tabelas de fatos, tabelas de dimensões, e hierarquias. O esquema estrela (Figura 6) inclui uma tabela central (tabela de fatos) contendo medidas quantitativas de natureza unitária ou transacional, a qual tem relacionamento com múltiplas tabelas dimensões.

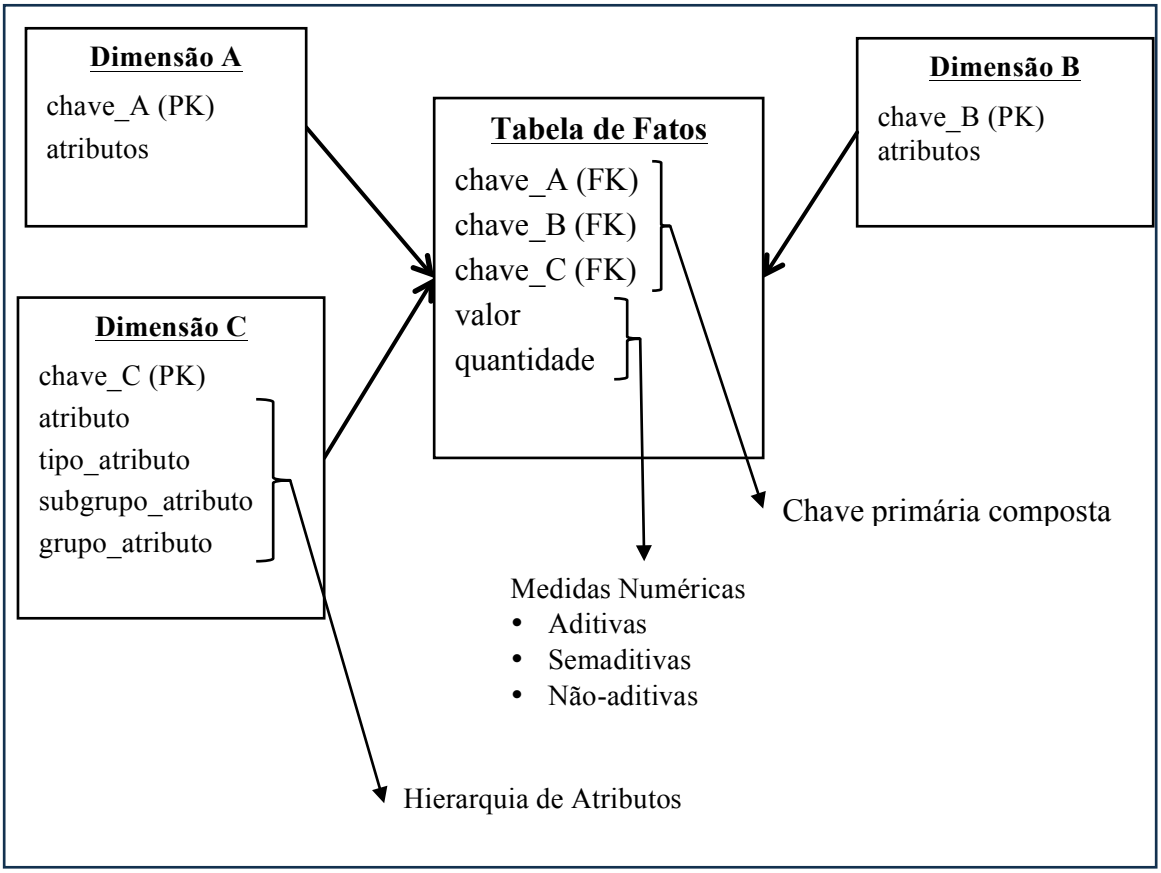


Figura 6 - Esquema estrela e suas principais características.

A **tabela de fato,** em um modelo multidimensional, é a tabela onde as medidas numéricas do negócio são armazenadas. O termo fato é utilizado para representar uma medida de negócio. Uma medida é uma tupla em uma tabela fato. A tabela de fatos possui sua própria chave primária identificadora, a qual é formada por um subconjunto de chaves estrangeiras.

As **medidas numéricas** em uma tabela de fatos podem ser classificadas em aditivas, semiaditivas e não aditivas. As medidas aditivas são obtidas por meio da soma de todas as suas dimensões. A medidas semiaditivas são obtida por meio da soma de algumas de suas dimensões. E as medidas não aditivas não podem ser somadas.

As **tabelas de dimensões** incluem os descritores textuais para a medida, essas tabelas contem muitas colunas ou atributos. Cada tabela dimensão é definida por uma única chave

primária. Cada atributo de dimensão pode consistir de um ou diversos níveis de agregação, sendo que esses níveis são organizados em hierarquias (PRAT; AKOKA; COMYN-WATTIAU, 2006).

Os atributos das tabelas de dimensões podem ser normalizados resultando em um esquema **floco de neve** fornecendo, dessa forma, suporte a hierarquias de atributos (GOPALKRISHNAN; KARLAPALEM, 1999). Os esquemas estrelas podem incluir múltiplos fatos conectados através dimensões em comum formando uma **constelação de fatos**. Essa estrutura permite que o usuário final percorra o esquema para realizar análises (LUJÁN-MORA; TRUJILLO; SONG, 2006).

3.5. OLAP

OLAP (*On-Line Analytical Processing*), processamento analítico de consultas, é uma categoria de tecnologia que apoia usuários de sistemas de tomada de decisões nas análises de dados do DW. A efetividade do OLAP sobre essas análises está relacionada com a capacidade de descrever e manipular dados de acordo com diferentes perspectivas ou dimensões (CABIBBO; TORLONE, 1998). Geralmente os dados do OLAP são apresentados como um cubo de dados multidimensional (GUPTA et al., 1997).

As operações OLAP incluem *roll-up*, *drill-drown*, *slice and dice*, *pivot* e *drill-across*. A operação *rollup* consiste em aumentar o nível de agregação ou reduzir o nível de detalhes na análise dos dados. A operação *drill-down* é o processo contrário da *rollup*, porque reduz o nível de agregação. A operação *slice and dice* permite que sejam restringidos os dados a uma determinada condição; o *slice* é a seleção através de uma ou mais dimensões para um determinado valor e o *dice* é uma projeção das dimensões através de faixas de valores. A operação *pivot* consiste na reorientação da visão multidimensional dos dados alterando a ordem das dimensões sobre a qual os dados são visualizados. A operação *drill-across,* através de no mínimo uma dimensão em comum, permite comparar medidas numéricas distintas (CIFERRI, 2002).

4. Trabalhos Correlatos

O levantamento e revisão bibliográfica realizados têm o objetivo de encontrar o estado da arte das implementações de ambiente DWing para o processamento de consultas analíticas em bases de dados de assistência a saúde para destacar as vantagens e desvantagens da utilização desse tipo de ambiente, comparado aos demais ambientes transacionais. Além disso, trabalhos correlatos a esta proposta foram analisados. A pesquisa bibliográfica envolveu também o estudo de ferramentas open-source que proporcionam a operabilidade de implementação e desenvolvimento do DW. Um comparativo das principais ferramentas estudadas contribuiu para a escolha da ferramenta a ser utilizada no projeto.

4.1. Implementações de DWing

Algumas publicações científicas foram estudadas sobre a aplicação dos conceitos de DWing a assistência em saúde, as quais analisam a implementação desse tipo de ambiente informacional que fornece consultas analíticas como suporte para a gestão de dados em saúde e no apoio a decisão.

Por meio do estudo das principais publicações sobre DWing aplicados a sistemas de atenção a saúde, podemos observar que no artigo de BANEK, TJOA e STOLBA (2006), o DW é utilizado como uma forma de armazenamento de uma grande quantidade de dados de saúde, funcionando mais especificamente como componente central de uma sistema que apoia a medicina baseada em evidência. O conceito de medicina baseada em evidência, quando aplicado ao processo de decisão clínica, acelera a transferência de descobertas de pesquisas clínica na prática levando a uma redução de custo para os pacientes e para as organizações de seguro de saúde, assim como para melhoria no processo de atenção a saúde como um todo. O artigo cita as contribuições desse estudo, que são: (1) o desenvolvimento de um modelo conceitual federado que integra com sucesso dados de baixo nível, de tempo segmentado e que mantém o nível de grão básico mais alto; (2) e a utilização de uma estrutura tipo XML para representar os dados de tempo segmentado, estendendo as abordagens existentes para junção de sistemas OLAP com

documentos XML. O estudo prossegue descrevendo como DW de diferentes organizações de seguros de saúde na Áustria são mesclados em um projeto de colaboração para a medicina baseada em evidência. O estudo de caso do DW federado é chamado de HEWAF (*Healthcare Warehouse Federation*) e é apresentado um modelo conceitual para as integrações serem feitas de forma coerente e com um mínimo de esforço. O modelo conceitual do HEWAF é baseado nos padrões internacionais de saúde HL7 (*Health Level Seven*) com a finalidade de alcançar um alto nível de generalização e portabilidade. Também foram utilizadas ontologias para fornecer a combinação de esquema entre a federação e os armazéns componentes. O estudo finaliza comentando sobre os principais desafios, os quais foram à integração dos dados de tempo segmentado e de baixa granularidade de administração de medicamentos no modelo conceitual padrão de DW multidimensional com medidas financeiras, e sobre o desenvolvimento de uma ontologia de DW domínio com a finalidade de automatizar o processo de combinação de esquema.

No artigo de KERKRI et al. (2001), é apresentado um *framework* de DW para reagrupar informações médicas de paciente de diferentes fontes de informações de saúde em um nível regional e integrá-los em um sistema de informação único mais abrangente. Os dados integrados nesse DW oferecem acesso a importantes dados a respeito do paciente, como por exemplo: seus fatores de risco e antecedentes. Assim, oferece melhorias em diagnósticos e em decisões médicas, fornecendo informações confiáveis, seguras e controladas para profissionais. O estudo propõe uma abordagem de integração de dados, a qual combina estratégias de DW com a metodologia de mediação de contexto. O *framework* de DWing de dados médicos chama-se *Epidware*. A metodologia desse framework é baseada em duas partes: o ambiente de DWing e a integração e interoperação de banco de dados. A primeira parte consiste no processo de implementação de DW, o qual inicia-se com a implementação dos *data marts* e segue com o desenvolvimento centralizado progressivo de dados estratégicos ou o desenvolvimento descentralizado de *data mart* de serviço específico. A segunda parte da metodologia de Epidware é a integração ou interoperação, a qual se inicia com a criação de um esquema-alvo que define a estrutura total dos dados para serem mesclados e, assim, criar o DW. Para modelar a estrutura de dados-alvo os projetistas devem estabelecer, com a ajuda de ferramentas ou da instituição médica, um

dicionário de descrição dos dados (metadados) e especificar as ferramentas para extração, tradução e integração dos dados.

A arquitetura do Epidware consiste de três componentes importantes que são: as fontes de informações, os empacotadores, e o integrador. As fontes de informações, no contexto médico, representam os bancos de dados ou qualquer sistema de informação utilizado em hospitais públicos ou privados, laboratórios de análises biomédicas, e departamentos de radiologia. Os empacotadores têm as funções de traduzir, monitorar e manter o anonimato. A tradução tem o papel de converter os dados de formatos de diversos sistemas de informação para um formato comum do DW, além de detectar alterações de dados relativos ao DW e propaga-las para o integrador, e por fim o anonimato tem o papel de codificar as informação obtidas das fontes para garantir confidencialidade dos dados do paciente. Por fim, o integrador tem a função de filtrar, refinar, integrar, e combinar as informações vindas das fontes locais.

EINBINDER et al. (2001) apresenta um estudo de caso de um repositório de dados que consiste em um DW com o intuito de fornecer funções para o suporte à pesquisa e educação em um centro médico acadêmico e também para prover dados para gestores e administradores. Nesse artigo foi criado o repositório de dados clínicos CDR (*Clinical Data Repository*) na Universidade de Virginia. O estudo de caso descreve esse sistema e discute a premissa de que um DW para um centro médico acadêmico também deve apoiar as missões acadêmicas da organização.

O estudo de caso do CDR apresenta a descrição do sistema: a arquitetura, sendo o CDR um DW relacional; o conteúdo, que consiste de dados de quatro fontes independentes; a interface de usuário, que executa em navegador Web e consiste de um dicionário de dados; a formulação de consulta, que permite que os usuários definam suas consultas de forma dirigida; a segurança, que consiste do estabelecimento de políticas para utilização e acesso do sistema.

O estudo prossegue com uma avaliação para entender necessidades de usuários, essa avaliação permite realizar melhorias no CDR. As abordagens utilizadas para avaliar o CDR incluem: monitoramento de população de usuário e padrões de uso; administração de uma pesquisa de usuário; rastreamento de consultas enviadas ao CDR e realização de entrevistas de acompanhamento por telefone.

O CDR mostra-se útil para os usuários administrativos em algumas situações, como por exemplo: quando dados clínicos são requeridos; quando os usuários precisam de dados

imediatamente; quando uma visão integrada dos dados é requerida; e quando um usuário tem uma questão operacional, mas não está conectado em fontes de dados administrativas típicas.

EVANS, LLOYD e PIERCE (2012) apresenta um estudo sobre um DW empresarial aplicado para os dados dos registros médicos eletrônicos (EMR - *Eletronic Medical Records*), o qual é denominado de DW empresarial EDW (*Enterprise DW*), e permite que todos os dados de uma organização com numerosas internações e unidade ambulatoriais sejam integrados e analisados. O estudo encontrou no EDW não apenas uma ferramental essencial para a gestão e tomada de decisão estratégica, mas também para o apoio de decisão clínica de determinado paciente.

O artigo descreve o *Intermountain Healthcare* que consiste em um sistema de atenção a saúde sem fim lucrativos que opera 22 hospitais e cerca de 179 clínicas, prontos-socorros, escritórios médicos e assistência em saúde residencial. Após apresentar o Intermountain Healthcare, o estudo traz as principais características do EDW, o qual contém fontes de dados e conjuntos de dados com todas as informações das áreas clínica e de negócio da organização. Os conjuntos de dados incluem dados de internações e clinico ambulatorial, dados financeiros de contas de paciente para gestão da cadeia de suprimento e gerenciamento de materiais, dados clínicos de laboratório para imagens, trabalho e entrega, cirúrgico, numerosos *data marts* apoiando centros, recursos humanos e folha de pagamento. O estudo apresenta um novo framework para coleção de dados e aplicações de apoio a decisão que demonstram o benefício de acesso aos dados de toda a empresa combinado com dados de admissão localizado no EMR. São apresentados também dois casos de uso de aplicações, os quais foram selecionados para demonstrar a funcionalidade e a capacidade do *framework*. Por fim, o estudo discute as principais características abordadas pelo EDW e o *framework* de apoio a decisão, o qual tem permitido a reutilização de métodos e *data marts* para reduzir tempo de desenvolvimento para novas aplicações. Uma limitação do sistema é que não tem uma base de dados empresarial contendo listas de pacientes atualizada diariamente. Os autores concluem afirmando que a utilização de dados no DW em organizações de saúde pode não somente fornecer dados precisos para gerenciamento e utilização de recursos, como também utilizar a informação no DW para melhorar cuidados ao paciente.

4.2. Ferramentas de Desenvolvimento de DWing

As ferramentas de implementação de DWing são utilizadas para extrair, transformar e carregar dados de diversas fontes legadas, assim como para estruturar e acessar os dados por meio de consultas OLAP. Estas ferramentas podem ser divididas em categorias de forma particular para cada etapa de implementação, tais como: categoria de processo ETL; categoria de servidor OLAP; categoria de cliente OLAP. As ferramentas integradas em conjunto são chamadas de BI *suites*.

Existem diversos tipos de licenças, as quais podem ser proprietárias, comerciais, *open source* (código aberto) e outras mais específicas. As licenças de ferramentas estudadas neste projeto de mestrado dizem respeito ao tipo *open source*. Estas ferramentas, independentemente da licença, podem ter duas categorias tais como: OLAP relacional (ROLAP) em que as tabelas do banco de dados são relacionais ou OLAP multidimensional (MOLAP) em que cubos multidimensionais são carregados.

Considerando as ferramentas por categoria e *open source*, existe uma variedade de ferramentas para cada processo de implementação do DWing. O processo ETL, pode ser implementado com as ferramentas Apatar, Clover.ETL, ETL Integrator, KETL, Kettle / Pentaho Data Integration, Octopus, Palo ETL Server, Pequel, Scriptella, ou Talend Open Studio / JasperETL. O processo de implementação dos servidores OLAP pode ser realizado com as ferramentas Mondrian / Pentaho Analysis Services ou Palo. A configuração de clientes OLAP pode ser realizada com as ferramentas FreeAnalysis, JPalo Client e JPalo Web Client, JMagallanes Olap & Reports, JPivot, JRubik, OpenI, e REX (THOMSEN; PEDERSEN, 2005).

Soluções de Business Intelligence (BI) têm apresentado abordagens e conjuntos de ferramentas eficientes para extração de informações de negócios espalhadas em diversos departamentos de uma organização (ZENG et al., 2006). As plataformas de BI geralmente satisfazem todas as etapas para o processo de implementação e de requisitos analíticos de um DWing fornecendo tecnologias para a modelagem dos dados, a carga dos dados, a apresentação analítica e elaboração estatística, a entrega e segurança dessas informações. Além disso, provêm consultas e relatórios, análises OLAP, painel de ferramentas e mineração de dados (GIOIA; CAZZIN; DAMIANI, 2008).

A literatura apresenta trabalhos comparativos entre as plataformas BI. Dentre eles podemos ressaltar os artigos de TERESO e BERNARDINO (2011), GOLFARELLI (2009) e THOMSEM e PEDERSEN (2005), os quais abordam as principais plataformas BI, tais como: PENTAHO (2014); SPAGOBI (2014); JASPERSOFT (2014); OpenI (2014); Palo (2014); e VANILLA (2014).

TERESO e BERNARDINO (2011) criaram uma tabela (Tabela 1) que apresenta um comparativo entre as ferramentas BI *open source*, considerando suas funcionalidades, características e sistema operativos.

Tabela 1 - Comparativo de plataformas BI *open source*

	SpagoBI	**OpenI**	**Pentaho**	**JasperSoft**	**Palo**	**Vanilla**
Funcionalidades						
Relatórios	✓	✓	✓	✓	✓	✓
Gráficos	✓	✓	✓	✓	✓	✓
Painel de Ferramentas	✓	✓	✓	✓	✓	✓
OLAP	✓	✓	✓	✓	✓	✓
ETL	✓	✗	✓	✓	✓	✓
Data Mining	✓	✓	✓	✗	✗	✓
KPIs	✓	✗	✓	✗	✗	✓
Exportação de dados	✓	✗	✓	✓	✓	✓
GEO/GIS	✓	✗	✗	✓	✗	✗
Consultas *ad hoc*	✓	✗	✓	✓	✓	✓
Características						
Licença GNU GPL	✓	✓	✓	✓	✓	✓
Versão Única	✓	✗	✗	✗	✓	✓

Sistemas Operacionais						
Linux	✔	✔	✔	✔	✔	✔
Windows	✔	✔	✔	✔	✔	✔
Unix	✔	✘	✔	✘	✘	✔

Como pode ser observado na Tabela 1, a ferramenta SpagoBI é a plataforma BI mais abrangente em funcionalidades em relação as outras plataformas *open source* apresentadas. A ferramenta OpenI é uma das ferramentas com menos recursos em comparação às demais, pois ela não fornece suporte aos processos ETL, bem como as funções de exibição de indicadores chave de desempenho (KPI - *Key Performance Indicators*), exportação de dados, suporte de localização GEO/GIS, e consultas *ad hoc*. O Pentaho é uma ferramenta bastante completa, porém tem uma funcionalidade a menos do que o SpagoBI, que é o suporte de localização GEO/GIS. O JasperSoft é uma ferramenta bem equilibrada em funcionalidades, mesmo não possuindo os processos de *Data Mining* e KPIs. A Palo também não fornece suporte a processos de *Data Mining* e KPIs, bem como funcionalidades de localização GEO/GIS, mas fornece outras funcionalidades o que a torna importante dentro do contexto BI. A plataforma Vanilla é semelhante ao Pentaho no que diz respeito às funcionalidades, sendo que um aspecto diferenciado para a plataforma Vanilla é que ela possui uma versão para dispositivos móveis (TERESO; BERNARDINO, 2011).

As plataformas BI apresentadas possuem licença GNU GPL (*General Public License*), que permite a distribuição de cópias, a alteração, e a utilização de partes dos softwares livres. Com exceção do SpagoBI, o qual é totalmente gratuita disponibilizando todas as suas funcionalidades, as demais ferramentas apresentadas têm duas versões: *Community* que dispõe parte das funcionalidades de maneira gratuita e *Enterprise* que é paga e incluí todas as funcionalidades da plataforma BI, bem como suporte e treinamento (GOLFARELLI, 2009).

Dentre as ferramentas estudas, o Pentaho foi escolhido porque oferece todos os módulos necessários para implementação do ambiente de DWing para o Mapa de Saúde, possui um amplo conteúdo bibliográfico para sua utilização (por exemplo livros, fóruns e sites), é um ferramenta

livre, e é bastante utilizado na literatura na área de informática aplicada a saúde de acordo com os estudos de LOPES (2014), PEREIRA (2014) e SOUZA e FREIRE (2014).

5. Desenvolvimento do Projeto

Considerando a complexidade das informações definidas pelo Mapa de Saúde e dos diversos sistemas de informações fontes de dados da saúde pública, nós definimos uma abordagem metodológica específica para a implementação do ambiente de DWing, a qual se ajusta ao contexto das informações da saúde pública.

A metodologia definida para esta implementação consiste basicamente de sete etapas, a saber: análise dos dados definidos pelo Mapa de Saúde; modelagem e implementação do banco de dados (BD) integrador; modelagem dimensional do DW; criação do processo ETL; configuração do servidor; análises; e testes.

5.1. Proposta de Arcabouço Computacional

O arcabouço computacional para DWing abrange um conjunto de conceitos e abstrações para a organização dos componentes e dos processos realizados em cada estágio desse domínio de conhecimento, que é o armazenamento de dados para análises em âmbito de tomadas de decisão. Portanto, para este projeto nós propomos um arcabouço computacional de DWing voltado para a saúde pública, ao qual chamamos de “Gestão Dinâmica baseada no Mapa de Saúde” (GDMS). A Figura 7 apresenta o arcabouço computacional GDMS proposto para o ambiente de DWing do Mapa de Saúde e seus principais componentes.

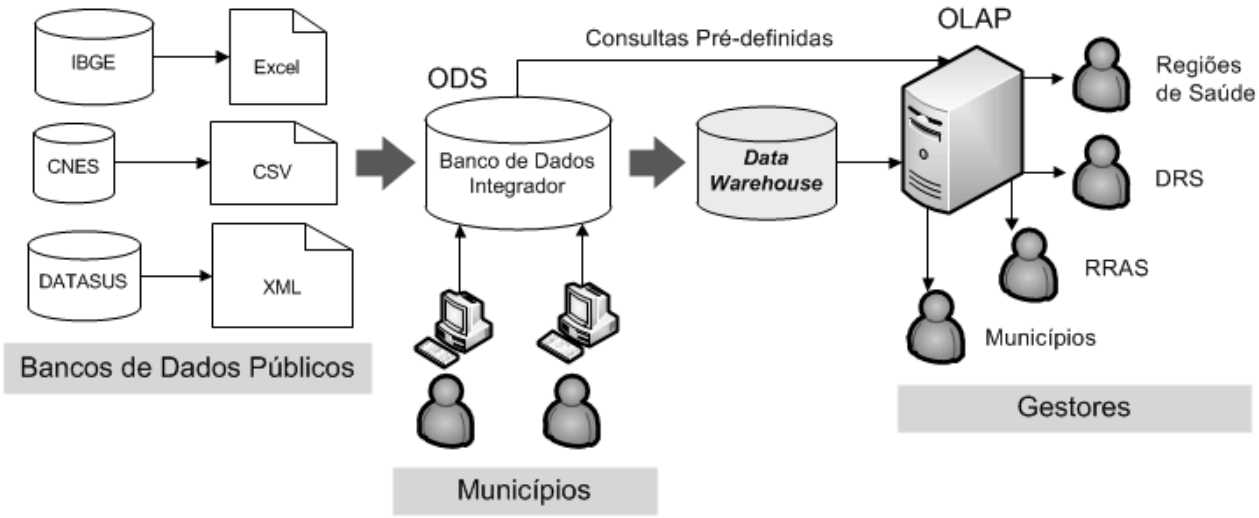


Figura 7 - Componentes do arcabouço computacional GDMS.

Como pode ser visto na Figura 7, o componente "Bancos de Dados públicos" do GDMS (BDpub-GDMS) representa os sistemas fontes de dados que estão definidos por meio dos documentos disponibilizados pelos municípios e definidos sobre o Mapa de Saúde, e também por outros sistemas públicos de dados em saúde, tais como o IBGE, o CNES, o DATASUS e outros.

Estes sistemas de informações fontes serão integrados para preparar os dados para o segundo componente, o ODS que corresponde ao "Banco de Dados Integrador" do GDMS (BDint-GDMS), onde serão resolvidos os problemas de inconsistências e conflitos de dados. O BDint-GDMS armazena os dados operacionais vindos dos sistemas fontes legados e sua base permite que os dados sejam atualizados. Por intermédio do BDint-GDMS serão implementadas consultas pré-definidas dos dados já integrados, com a finalidade de se estabelecer um ambiente estático de consultas ao Mapa.

O componente seguinte é o "*Data Warehouse*" do GDMS (DW-GDMS), o qual tem como papel armazenar os dados multidimensionais. Este componente contém os esquemas estrelas, os esquemas floco de neve, as constelações de fatos e as medidas de interesse sobre as informações oriundas do BDint-GDMS, as tabelas de fatos e suas respectivas tabelas de dimensões. Desta forma, os dados multidimensionais contidos nesta base proverão ao ambiente o suporte para a realização de consultas analíticas.

Por fim, o último componente é o servidor OLAP (OLAP-GDMS), com o qual os gestores poderão executar suas operações analíticas sobre os esquemas multidimensionais construídos no DW-GDMS, por exemplo, as operações *rollup*, *drill-down*, *slice-and-dice*, *pivot*, e *drill-across*.

5.2. Etapas do Processo de Desenvolvimento

A Figura 8 apresenta a sequência das etapas que estão sendo seguidas para a construção do arcabouço computacional GDMS.

Fase 1 - Análise dos dados definidos pelo Mapa de Saúde e disponibilizados pelas fontes públicas. Esta é a etapa inicial do processo de desenvolvimento, a qual consiste na análise de todas as informações contidas nos diversos sistemas públicos fontes de saúde, tais como o IBGE, o CNES, o DATASUS e outros. As informações de saúde nestes sistemas estão disponíveis em formatos variados, como por exemplo em XML (*extensible markup language*), em Microsoft Excel, em CSV (*comma-separated values*), HTML (*hypertext markup language*), e outros formatos conhecidos. Sendo assim, esta fase exige a limpeza, formatação e padronização dessas informações com a finalidade de posteriormente serem integradas no BD integrador.

Fase 2 - Modelagem e implementação do BDint-GDMS. Esta segunda etapa representa a modelagem de todas as informações obtidas na etapa inicial do processo de implementação. Dessa forma, os dados serão integrados e padronizados para um formato comum.

Fase 3 - Modelagem dimensional do DW-GDMS. Esta etapa representa a modelagem multidimensional dos dados provenientes do banco de dados integrador. Na modelagem dimensional serão construídos os esquemas estrelas por meio da definição de suas medidas de interesse. Assim, serão construídas tabelas de fatos, tabelas de dimensões e suas hierarquias de atributos.

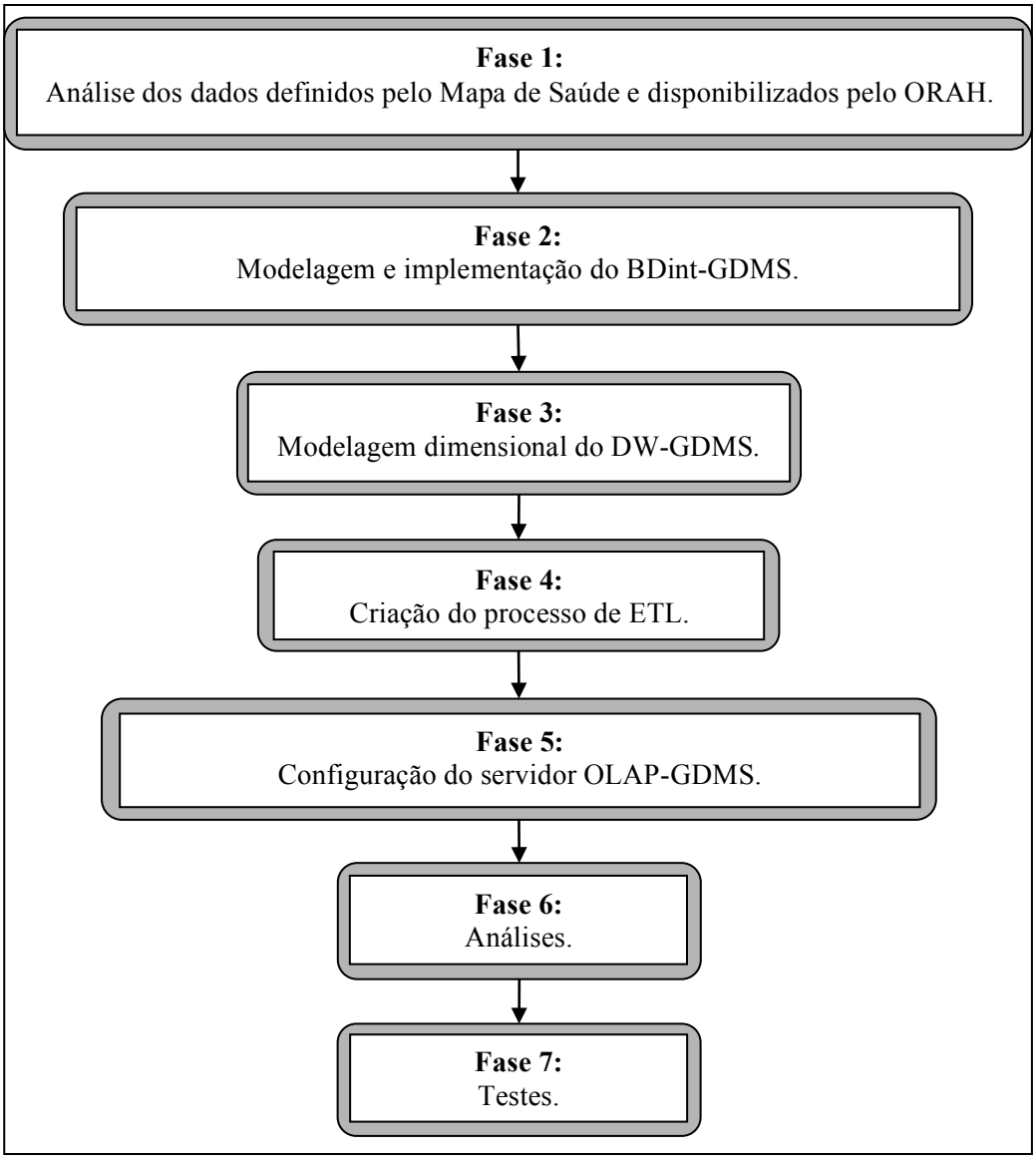


Figura 8 - Etapas da implementação do ambiente de Data warehousing

Fase 4 - Geração do processo de ETL. Nesta fase serão configurados os processos de extração dos dados dos sistemas fontes de saúde, o processo de transformação dos dados com o propósito de fornecer uma visão consistente e padronizada da informação, e o processo de carga dos dados a fim de alimentar o DW-GDMS com os dados já processados.

Fase 5 - Configuração do servidor OLAP-GDMS. Nessa fase, será configurado o servidor com as operações OLAP de acordo com as necessidades dos gestores. Serão construídos os cubos

multidimensionais que permitirão a criação ad-hoc de análises e a navegação nas hierarquias dimensionais.

Fase 6 - Geração de análises. Nesta etapa são construídas as interfaces de análise e as consultas sobre os dados contidos no DW-GDMS, bem como as consultas *ad hoc* complexas.

Fase 7 - Testes. Nesta última etapa, serão realizados testes junto aos gestores de saúde e potenciais usuários do sistema, com a finalidade de avaliar a usabilidade do ferramental e verificar se os relatórios e análises gerados pelo DW-GDMS estão corretos e se estão de acordo com os dados de saúde fornecidos.

5.3. Ferramentas de Desenvolvimento

Para o desenvolvimento deste projeto foram utilizadas as seguintes ferramentas:

- *MySQL Workbench* 5.1 para a modelagem do DW;
- *PostgreSQL* como sistema gerenciador de banco de dados relacional (POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2014);
- *Pentaho BI Suite* (PENTAHO, 2014), plataforma utilizada para a criação do ambiente de DWing*;*
- *Pentaho BI Platform*, que pertence ao conjunto de ferramentas do *Pentaho BI Suite* e foi utilizada para a visualização e configuração da informação contida no DWing;
- *Pentaho Data Integration (Kettle)* (CASTERS; BOUMAN; DONGEN, 2010), que também pertence ao conjunto de ferramentas *Pentaho BI Suite* e tem a função de dar suporte ao processo ETL.
- *Pentaho Data Analisys (Mondrian)* (HYDE, 2014), utilizada como servidor OLAP.

O *Pentaho BI Suite* é uma coleção de programas computacionais que trabalham juntos para criar e entregar soluções para os problemas do mundo real. Este conjunto de programas oferece diversos recursos que disponibilizam funcionalidades muito básicas, tais como autenticação de usuário ou gerenciamento de conexão de banco de dados. Outros recursos oferecem funcionalidades em alto nível, tais como visualização de dados utilizando gráficos e tabelas. Esse conjunto de ferramentas pode ser visto como uma pilha de componentes, que a cada

nível traz algumas funcionalidades. As principais camadas desta pilha são identificadas como a camada de apresentação, localizada no topo da pilha e a camada de integração localizada na base da pilha, bem como a camada do meio da pilha, a qual traz como funcionalidades os relatórios, as análises, os painéis ferramentais e o gerenciamento de processos (BOUMAN; DONGEN, 2009).

A Figura 9 apresenta os componentes que compõe a pilha de funcionalidades da plataforma *Pentaho BI Suite*.

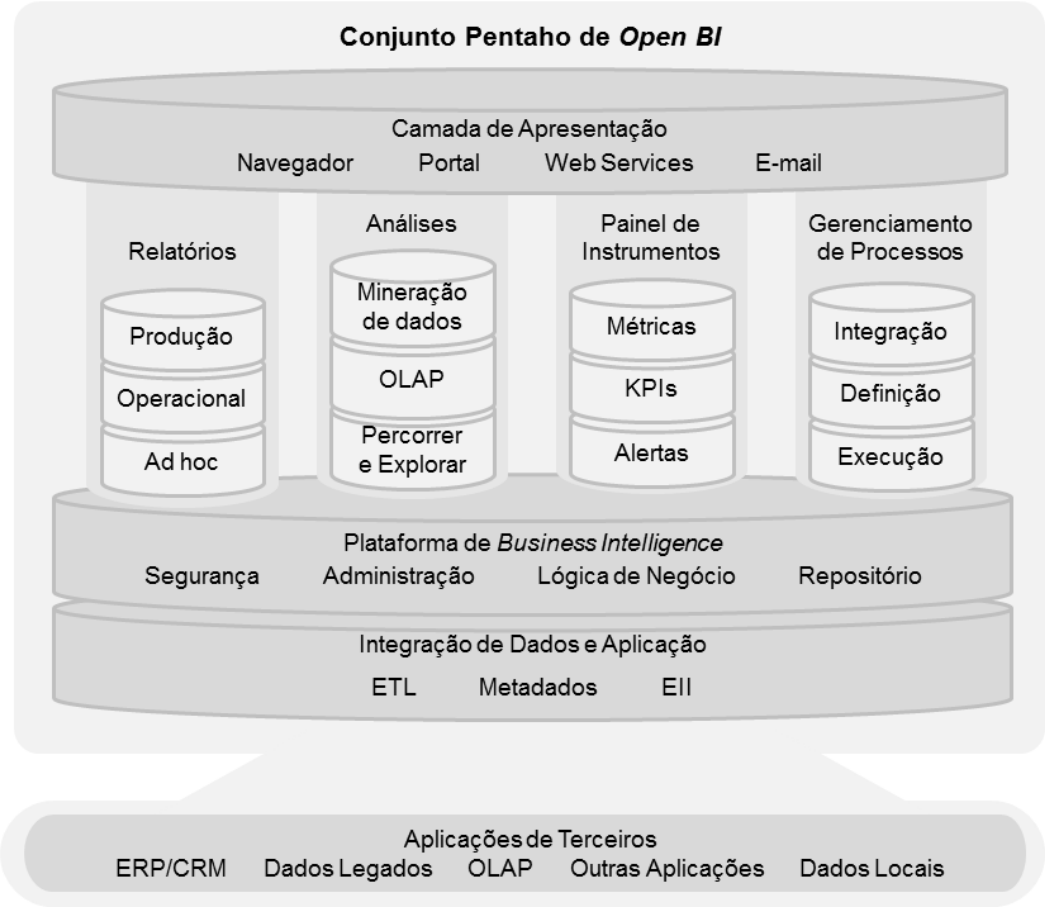


Figura 9 – Diagrama de Componentes do *Pentaho BI Suite* traduzido e adaptado de BOUMAN e DONGEN (BOUMAN; DONGEN, 2009).

6. Resultados

A proposta deste projeto teve como objetivo desenvolver o arcabouço computacional GDMS, para tomadas de decisões por meio dos indicadores do Mapa de Saúde, com uma visão mais ampla e abrangente das informações. Todas as etapas do processo de desenvolvimento inicialmente previstas foram concluídas e os resultados estão detalhados e explicados nas subseções subsequentes.

6.1. Análise dos dados definidos pelo Mapa de Saúde

A primeira fase do processo de desenvolvimento foi o estudo dos documentos do Mapa de Saúde que foram disponibilizados pelo ORAH. O Mapa de Saúde define os indicadores de interesse nas análises sobre os dados de saúde segundo divisão geográfica. Portanto, os indicadores foram analisados e compreendidos para melhor adaptarmos o GDMS às necessidades especificadas no Mapa de Saúde.

Os documentos do Mapa de Saúde que foram estudados são arquivos, no formato Microsoft Word, pertencentes a RRAS13, eles estão organizados por Regiões de Saúde e divididos em capítulos, seções e subseções. Estes documentos, atualmente, foram disponibilizados em uma página Web pelo ORAH, onde é possível selecionar a região de saúde de interesse e navegar por seus capítulos por meio de um menu.

Assim, nós realizamos um estudo sobre esta página do ORAH, acessamos todos os capítulos, seções e subseções, verificamos todas as tabelas e suas fontes (por exemplo: IBGE, CNES, SES, SEADE, SIH, SINAN, etc.) e construímos uma modelagem de um banco de dados baseada nestas tabelas do Mapa de Saúde.

6.2. Modelagem do Banco de Dados Integrador

Na segunda fase nós realizamos a modelagem do BDint-GDMS, onde foi necessário um estudo minucioso de todas as informações contidas nos documentos do Mapa de Saúde. Como auxílio para essa etapa, todas as informações concernentes foram mapeadas em uma planilha, a

qual descreve de forma detalhada a qual sistema fonte o dado pertence, qual o formato de origem, qual sua granularidade, e em qual capítulo, seção, ou subseção dos documentos do Mapa de Saúde esse dado está descrito.

A Figura 10 apresenta uma visão parcial do Diagrama Entidade-Relacionamento (DER) da modelagem do BDint-GDMS, para organização das informações oriundas das bases de dados legadas. O BDint-GDMS possui 38 tabelas e o total de 702 atributos (em média 18,47 atributos por tabela).

Os dados modelados de acordo com o DER apresentado na Figura 10 referem-se às informações oriundas dos sistemas fontes de saúde que estão definidas pelo Mapa de Saúde. Estas informações são disponibilizadas de forma estática em diversos formatos. Os dados das diversas fontes heterogêneas da saúde são integrados de forma a mantê-los formatados, padronizados e combinados para serem posteriormente carregados no DW-GDMS. Além disso, também realizamos o mapeamento das informações dos BDpub-GDMS → BDint-GDMS, com a intenção de mantermos alguns metadados para o desenvolvimento e carga dos dados entre estas duas bases de dados. A Tabela 2 apresenta parcialmente o mapeamento das informações dos BD públicos para o BD integrador.

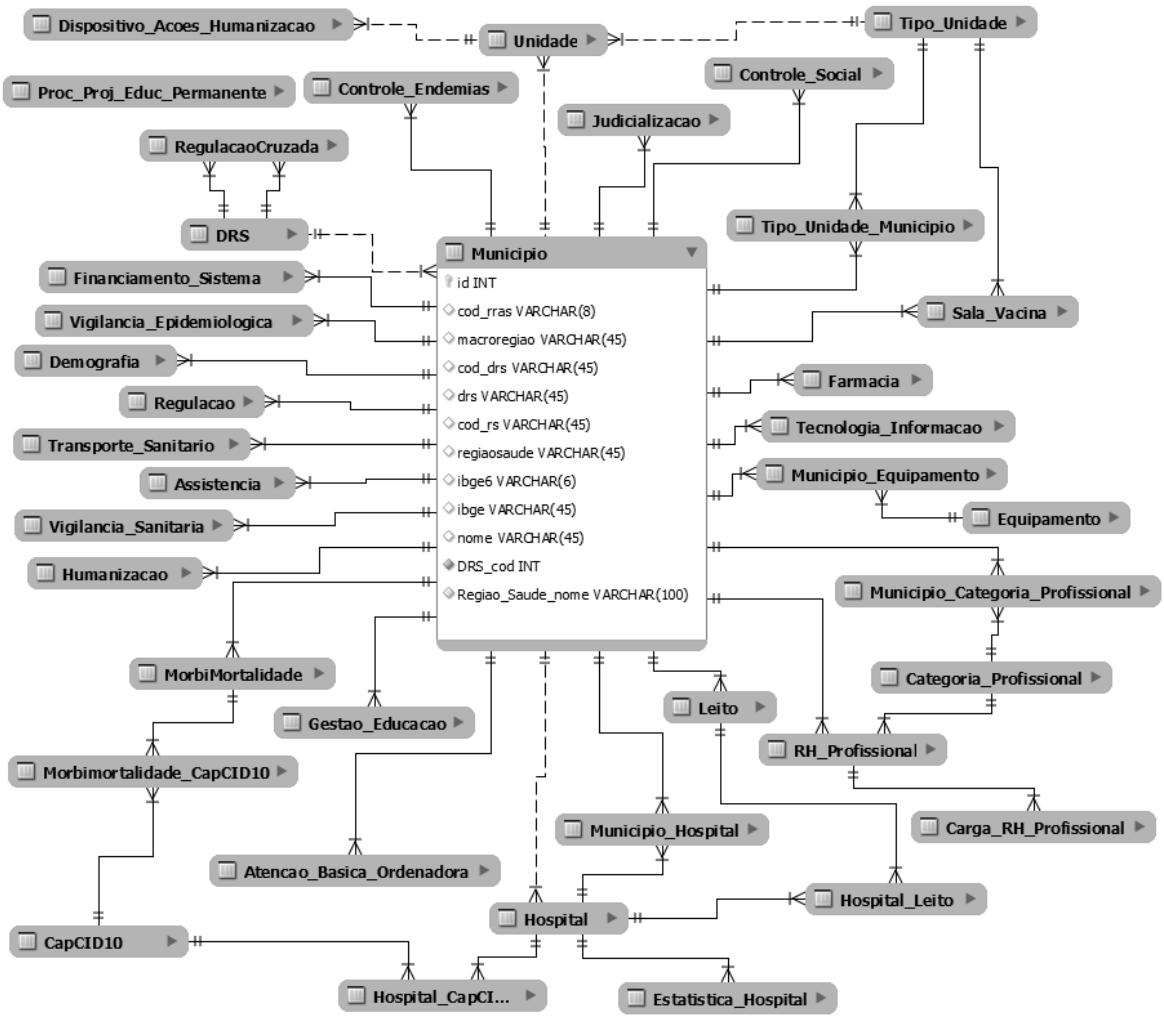


Figura 10 – DER parcial do BD integrador do GDMS.

Como podemos visualizar na tabela 2, o mapeamento contêm os atributos obtidos pela extração dos dados dos BDpub-GDMS. Neste mapeamento são especificados os tipos de arquivos nos quais as informações estão disponíveis. Além disso, também é especificado o capítulo, a seção e/ou subseção e qual o nível de detalhe destas informações no Mapa de Saúde podendo, por exemplo, serem detalhadas por município ou agregadas por DRS, Região de Saúde ou RRAS. Para todos esses dados extraídos dos BDpub-GDMS foram criadas relações e atributos no BDint-GDMS.

Tabela 2 - Mapeamento de informações dos BDpub-GDMS → BDint-GDMS

Tipo Arquivo	Fonte	Relação	Atributo	Capítulo	Seção	Grão
xls, csv	SES/SEADE	MorbiMortalidade	obitos_menor1	2	1	Municípios
xls, csv	SES/SEADE	MorbiMortalidade	nascidos_vivos	2	1	Municípios
xls, csv	SES/SEADE	MorbiMortalidade	obitos_neonatal	2	2	Municípios
xls, csv	SES/SEADE	MorbiMortalidade	obitos_posneonatal	2	3	Municípios
word/tabnet	DATASUS/ MS	MorbiMortalidade	obitos_maternos	2	4	Municípios
Word	Não informado	MorbiMortalidade	num_obitos	2	5	Região de Saúde, RRAS
xls, csv	SES/SEADE	Morbimortalidade_ CapCID10	num_obitos_masc	2	6	Municípios, DRS
xls, csv	SES/SEADE	Morbimortalidade_ CapCID10	num_obitos_fem	2	6	Municípios, DRS
xls, csv	SES/SEADE	Morbimortalidade_ CapCID10	obitos_causa_mal_ definida	2	7	Municípios
xls, csv	SES/SEADE	Morbimortalidade_ CapCID10	obitos_total	2	7	Municípios

6.3. Modelagem Dimensional do DW

A partir do BDint-GDMS modelado e implementado no PostgreSQL nós realizamos a modelagem do DW-GDMS com base nos dados mapeados neste repositório integrado. No total foram criados 27 esquemas estrelas do DW-GDMS, cada esquema com sua respectiva tabela de fatos, cada esquema possui pelo menos 2 dimensões. Em termos de capítulos do Mapa de Saúde, alguns capítulos contêm um esquema estrela e outros contêm mais de um esquema estrela. Nas subseções subsequentes serão apresentados os diagramas de alguns dos esquemas estrelas desenvolvidos para o repositório de DW-GDMS.

Inicialmente, a tabela de "Morbimortalidade" do BDint-GDMS foi utilizada como ponto de partida para a construção dos esquemas estrelas, e através dela nós realizamos a modelagem de quatro esquemas estrelas, a saber: o esquema estrela de "Mortalidade", "Morbidade", "Mortalidade Segundo Causas Específicas" e "Amostras de Água".

6.3.1. Esquema Estrela de Mortalidade

O esquema estrela de "Mortalidade" contempla uma tabela de fatos com as medidas numéricas de óbitos por local de ocorrência e óbitos por local de residência, as dimensões criadas para este esquema são: a dimensão ano, a dimensão município, a dimensão faixa etária, a dimensão sexo e a dimensão capítulos do CID10. A Figura 11 mostra o DER do esquema estrela de mortalidade.

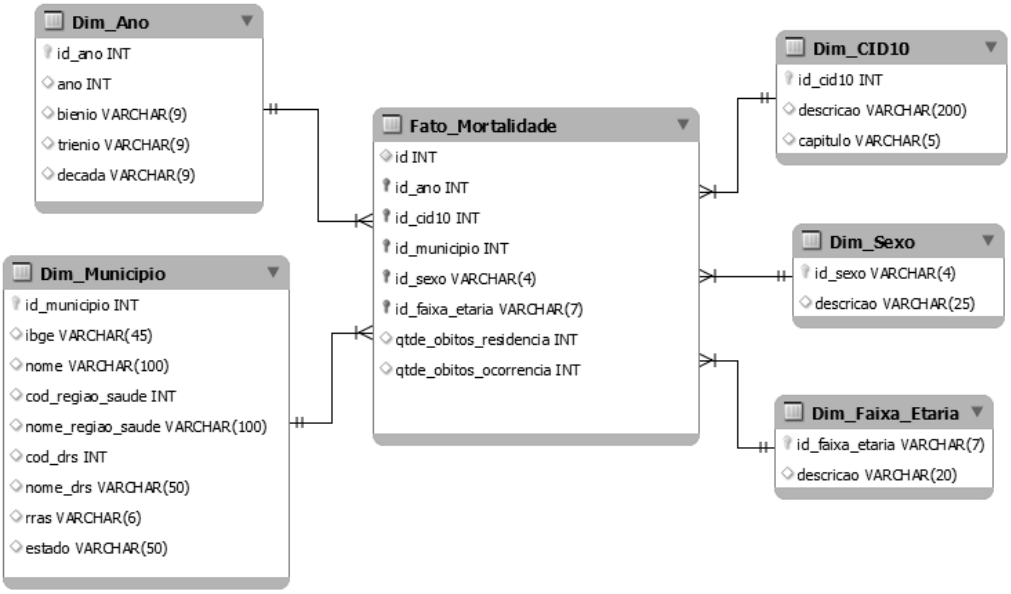


Figura 11 - DER do Esquema Estrela de Mortalidade.

Como podemos observar no esquema estrela de mortalidade, as dimensões ano e município contêm respectivamente uma hierarquia entre seus atributos, por exemplo, os atributos "estado → rras → nome_drs → nome_regiao_saude → nome" expressos na dimensão município demonstram uma hierarquia entre eles, sendo o atributo "estado" o nível mais agregado de um dado e o atributo "nome" (nome do município) o nível mais detalhado de um dado. Semelhantemente, ocorre na dimensão ano onde os atributos "década → biênio → ano" apresentam uma hierarquia, sendo que o atributo "década" representa o dado mais agregado e o atributo "ano" representa o dado mais detalhado. Portanto, as granularidades das dimensões "Dim_Municipio" e "Dim_Ano" são representadas, respectivamente, pelo nome do município e pelo ano de ocorrência do fato.

6.3.2. Esquema Estrela de Morbidade

O segundo esquema estrela modelado para o DW é o de “Morbidade”, ele contempla uma tabela de fatos cujas medidas numéricas são a quantidade de internações por local de internação e quantidade de internações por local de ocorrência. Este esquema se relaciona com as mesmas dimensões do esquema de “Mortalidade”, isto é, as dimensões ano, município, capítulos do CID10, sexo e faixa etária são compartilhadas com este esquema. A Figura 12 apresenta o DER do esquema estrela de “Morbidade”.

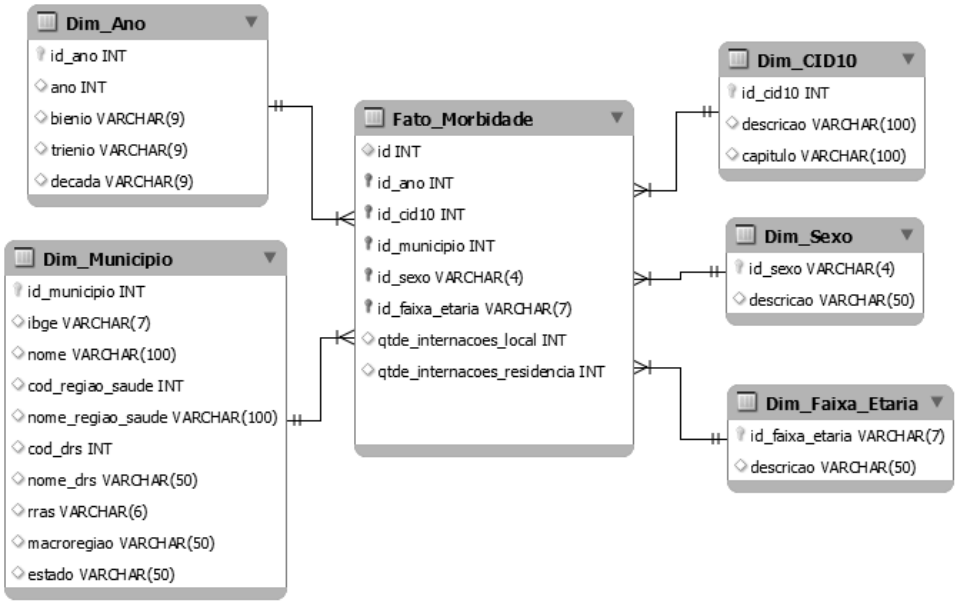


Figura 12 - DER do Esquema Estrela de Morbidade.

A tabela de fatos “Fato_Morbidade” se relaciona com as mesmas dimensões que a tabela de fatos “Fato_Mortalidade”. Portanto, os esquemas compartilham suas dimensões formando um esquema mais abrangente denominado constelação de fatos, que será explicado em detalhes posteriormente.

6.3.3. Esquema Estrela de Mortalidade por Causas Específicas

O terceiro esquema estrela modelado é o de “Mortalidade por Causas Específicas”, ele foi criado visando contemplar as mortalidades segundo algumas causas levantadas pelos documentos do Mapa de Saúde, ele possui como medidas numéricas a quantidade de óbitos e a quantidade de nascidos vivos. Análogo aos esquemas anteriores, as dimensões município e ano também são compartilhadas por este esquema, com uma nova dimensão, chamada de “Dim_Natureza”, que determina as causas de interesse especificadas no Mapa de Saúde. A Figura 13 apresenta o DER do esquema de “Mortalidade por Causas Específicas”.

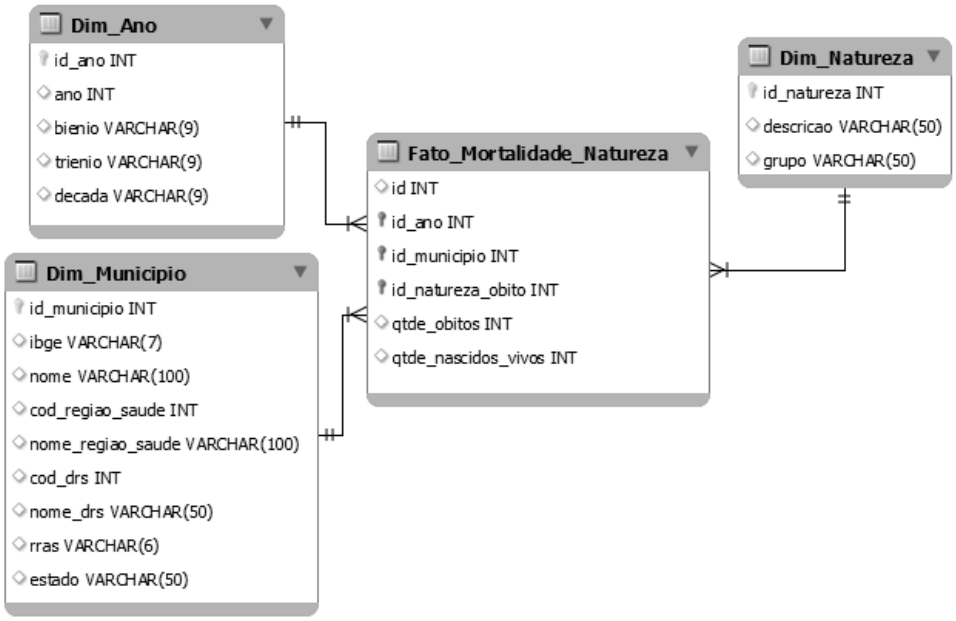


Figura 13 - DER do Esquema Estrela de Mortalidade por Causas Específicas.

A dimensão “Dim_Natureza_Obito” apresenta as causas levantadas pelos documentos do Mapa de Saúde e, como pode ser visto, nesta dimensão existe uma hierarquia entre seus atributos “grupo_natureza → natureza” sendo que o primeiro atributo representa o dado mais agregado e o segundo atributo o dado mais detalhado.

6.3.4. Esquema Estrela de Amostra de Água

O quarto esquema estrela implementado foi o de "Amostra de Água", este esquema também compartilha as dimensões ano e município e contém uma dimensão chamada "Dim_Tipo_Amostra" cujos dados representam o tipo de amostra que está sendo analisada. A tabela de fatos, denominada "Fato_Amostra", contém duas medidas numéricas aditivas – "qtde_realizadas" e "qtde_satisfatorias", as quais representam a quantidade de amostras realizadas e a quantidade de amostras satisfatórias. A Figura 14 apresenta o DER do esquema estrela de amostra de água.

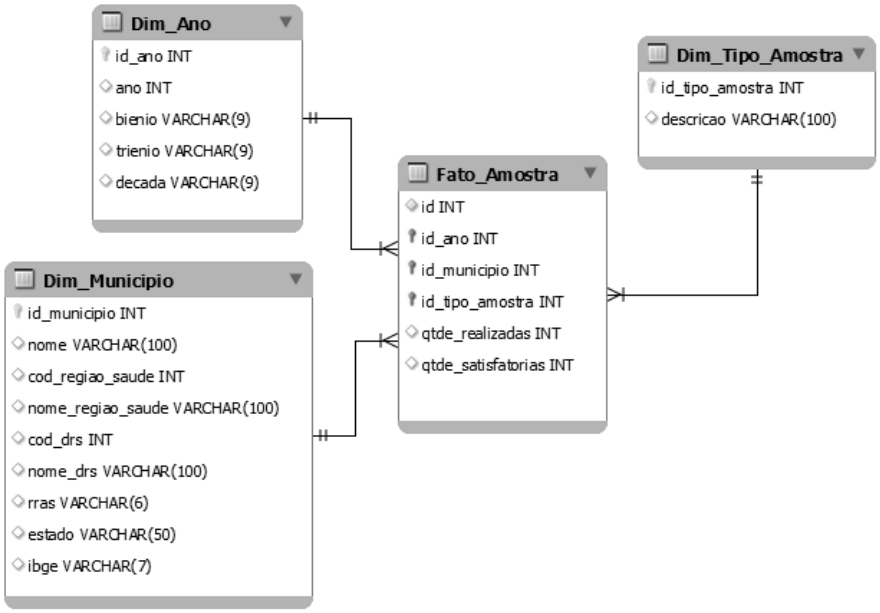


Figura 14 - DER do Esquema Estrela de Amostra de Água.

6.3.5. Esquema Estrela de Equipamento Hospitalar

O quinto esquema estrela criado foi o "Equipamento Hospitalar", este esquema contém uma tabela de fatos chamada de "Fato_Equipamento" cujas medidas numéricas são "qtde_existente" e "qtde_em_uso", as quais representam, respectivamente, a quantidade de equipamentos existentes e a quantidade de equipamentos em uso. O esquema "Equipamento

Hospitalar" também compartilha a dimensão "Dim_Ano", entretanto, diferentemente dos esquemas anteriores, ao invés de compartilhar a dimensão "Dim_Municipio" a tabela de fatos deste esquema relaciona-se com a dimensão "Dim_Hospital", consequentemente, os dados deste esquema estão disponíveis segundo hospitais. Além disso, a dimensão "Dim_Hospital" relaciona-se com a dimensão "Dim_Municipio", permitindo, deste modo, identificar a qual município os hospitais pertencem.

Ademais, a tabela "Fato_Equipamento" relaciona-se com uma dimensão chamada "Dim_SUS", a qual determina se as medidas são por SUS (Sistema Único de Saúde) ou pelo sistema privado. A Figura 15 apresenta o DER do esquema estrela de equipamento hospitalar.

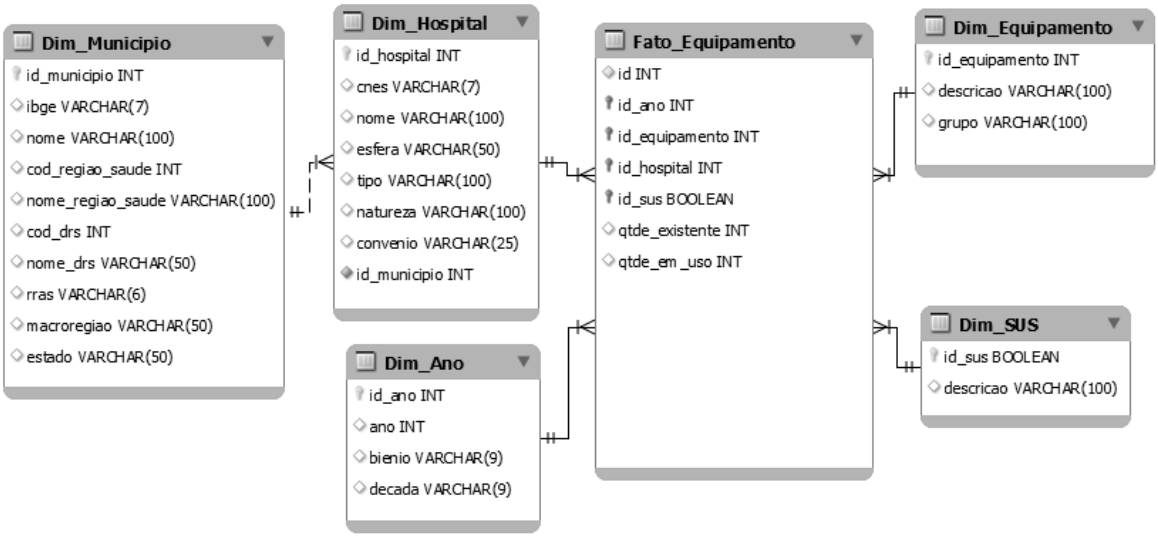


Figura 15 - DER do Esquema Estrela de Equipamento Hospitalar.

6.3.6. Esquema Estrela de Leito Hospitalar

O sexto esquema implementado é o esquema estrela de "Leito Hospitalar", este esquema compartilha as mesmas dimensões que o esquema estrela de equipamento hospitalar, tais como as dimensões "Dim_Hospital", "Dim_Ano" e "Dim_SUS". Contudo, no lugar da dimensão "Dim_Equipamento" foi implementado uma dimensão chamada "Dim_Especialidade" para

registrar os dados de especialidades dos leitos hospitalares. A Figura 16 apresenta o DER do esquema estrela de leito hospitalar.

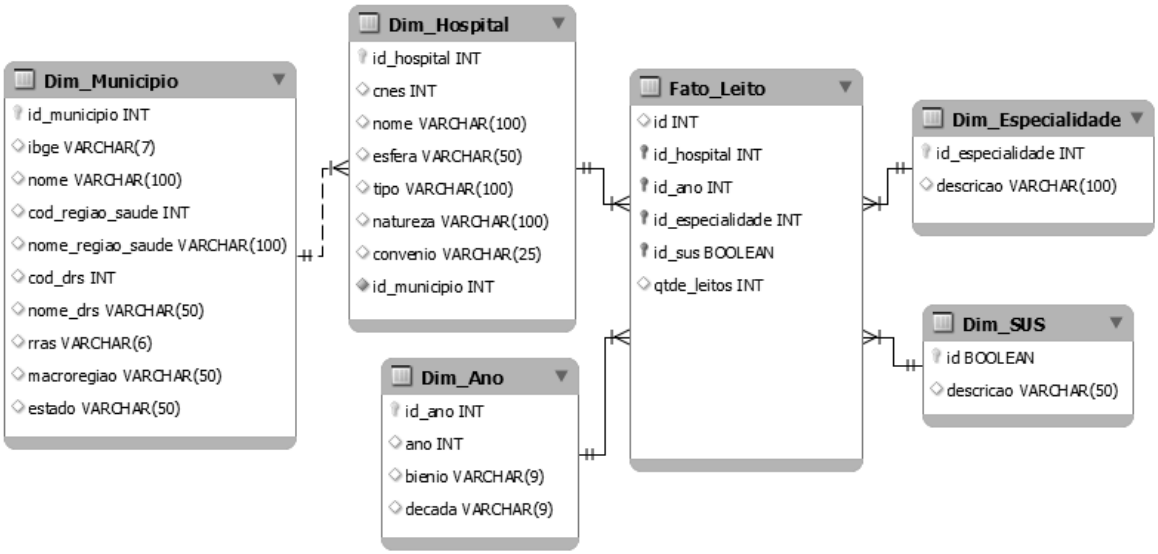


Figura 16 - DER do Esquema Estrela de Leito Hospitalar.

Como podemos observar na Figura 16, a tabela de fatos deste esquema é chamada de "Fato_Leito" cuja medida é a "qtde_leitos", a qual representa a quantidade de leitos que os hospitais possuem.

6.3.7. Esquema Estrela de Análise Profissional

Por fim, o último esquema que será apresentado aqui neste projeto é o esquema de "Análise Profissional". Semelhantemente aos esquemas de equipamento hospitalar e o de leito hospitalar, este esquema também contém as dimensões "Dim_Hospital", "Dim_Ano", "Dim_SUS" e "Dim_Municipio" relacionando-se com "Dim_Hospital". A Figura 17 apresenta o DER do esquema estrela de "Análise Profissional".

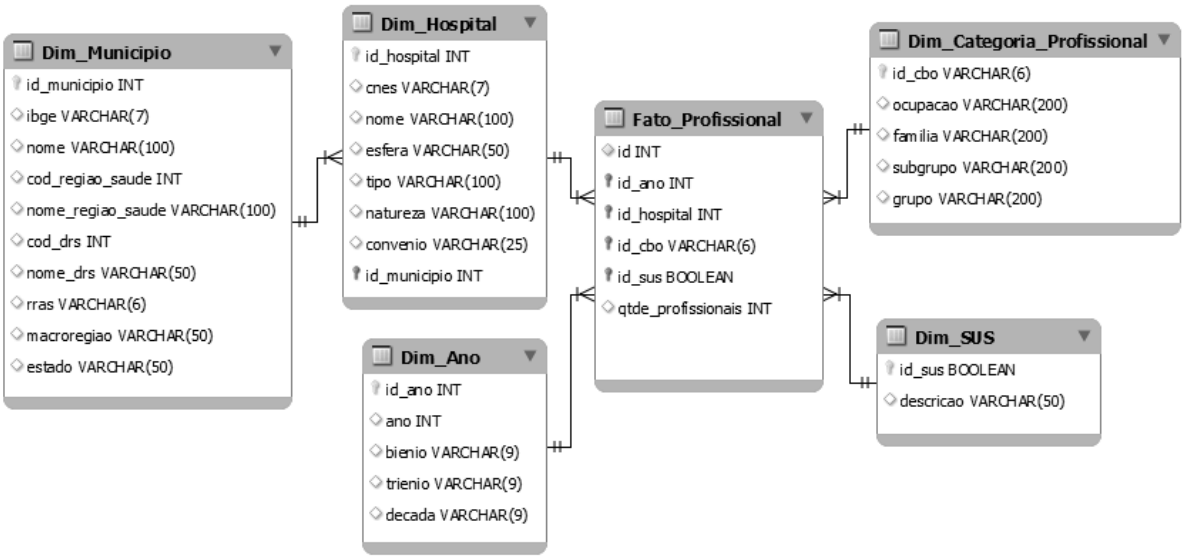


Figura 17 - DER do Esquema Estrela de Análise Profissional**.**

O esquema estrela de "Análise Profissional" contém uma dimensão chamada de "Dim_Categoria_Profissional", que representa os dados de ocupação profissional de acordo com a Classificação Brasileira de Ocupações – CBO (CLASSIFICAÇÃO BRASILEIRA DE OCUPAÇÕES, 2015). A tabela de fatos "Fato_Profissional" contém a medida numérica aditiva "qtde_profissionais" que representa a quantidade de profissionais.

6.3.8. Constelação de Fatos

Todos os esquemas estrelas descritos anteriormente contêm dimensões que são comuns entre eles, ou seja, dimensões que são compartilhadas entre suas respectivas tabelas de fatos. Como já foi citado anteriormente as dimensões compartilhadas são "Dim_Ano", "Dim_Municipio", "Dim_Hospital", "Dim_SUS", "Dim_CID10", "Dim_Sexo" e "Dim_Faixa_Etaria".

A dimensão "Dim_Ano" é compartilhada por todos os esquemas estrelas que fazem parte do DW-GDMS. A dimensão "Dim_Municipio", nos exemplos apresentados neste projeto, é compartilhada entre as tabelas fatos dos esquemas estrelas de "Mortalidade", de "Morbidade", de "Mortalidade por Causas Específicas" e de "Amostra de Água". As tabelas de fatos dos demais

esquemas estrelas, tais como o de “Equipamento Hospitalar”, de “Leito Hospitalar” e de “Análise Profissional” compartilham a dimensão “Dim_Hospital” e a dimensão “Dim_SUS”. Particularmente, a dimensão “Dim_CID10” é compartilhada entre as tabelas de fatos dos esquemas estrelas de “Mortalidade” e de “Morbidade”. Já as dimensões “Dim_Sexo” e “Dim_Faixa_Etaria” além dos esquemas estrelas de “Mortalidade” e de “Morbidade” elas também são compartilhadas pelo esquema estrela de “Demografia”.

Portanto, pode-se afirmar que por meio destes esquemas estrelas é possível formar de uma a quatro constelações de fatos. A primeira constelação é dada pela dimensão “Dim_Ano”, que é compartilhada entre todos os esquemas estrelas modelados no DW-GDMS, deste modo, por meio desta dimensão podemos formar uma ampla constelação de fatos por ano. A segunda constelação é dada pela dimensão “Dim_Municipio”, que também é compartilhada pelos esquemas estrelas de todo o DW-GDMS. A terceira constelação é dada pela dimensão “Dim_Hospital”, que é compartilhada pelos esquemas estrelas de “Equipamento Hospitalar”, de “Leito Hospitalar” e de “Análise Profissional”. A quarta constelação é dada pela dimensão “Dim_SUS” que também é compartilhada pelos esquemas estrelas de “Equipamento Hospitalar”, de “Leito Hospitalar” e de “Análise Profissional”. A quinta constelação é dada pelas dimensões “Dim_CID10”, “Dim_Sexo” e “Dim_Faixa_Etaria”, que são compartilhadas pelos esquemas estrelas de “Mortalidade” e “Morbidade”.

A título de exemplificação, somente as constelações de fatos dadas pelas dimensões “Dim_Municipio” e “Dim_Hospital” serão apresentadas. A Figura 18 mostra o DER da constelação de fatos por municípios.

A dimensão “Dim_Municipio” relaciona-se com quatro tabelas de fatos, a saber: “Fato_Mortalidade”, “Fato_Morbidade”, “Fato_Mortalidade_Natureza’ e “Fato_Amostra”. O relacionamento entre estas tabelas de fatos e a “Dim_Municipio” forma um esquema mais amplo que é a constelação de fatos por municípios.

Semelhantemente ocorre na constelação de fatos por hospitais (Figura 19), na qual a dimensão “Dim_Hospital” é compartilhada entre três tabelas de fatos, a saber: “Fato_Equipamento”, “Fato_Leito” e “Fato_Profissional”.

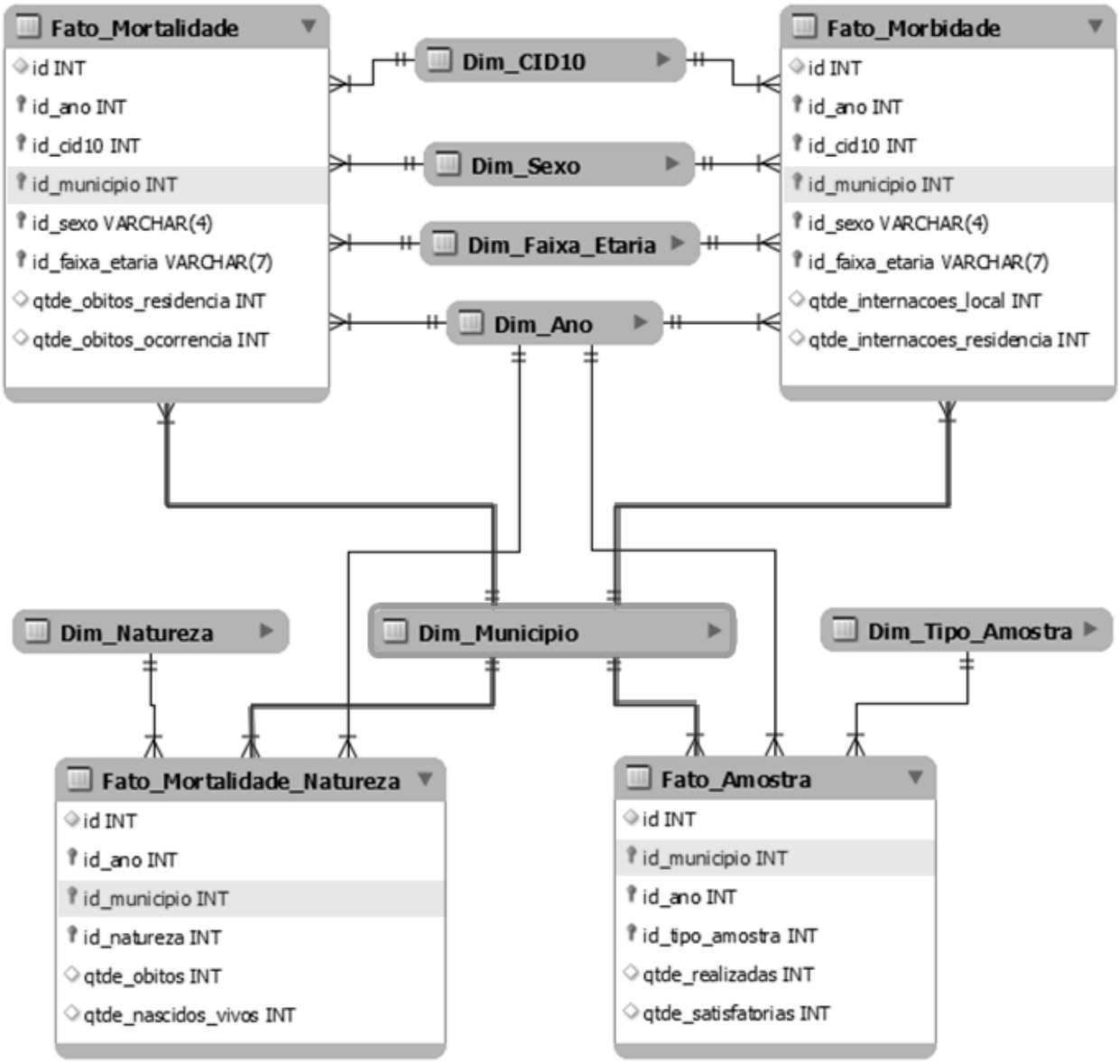


Figura 18 – DER da Constelação de Fatos por Municípios.

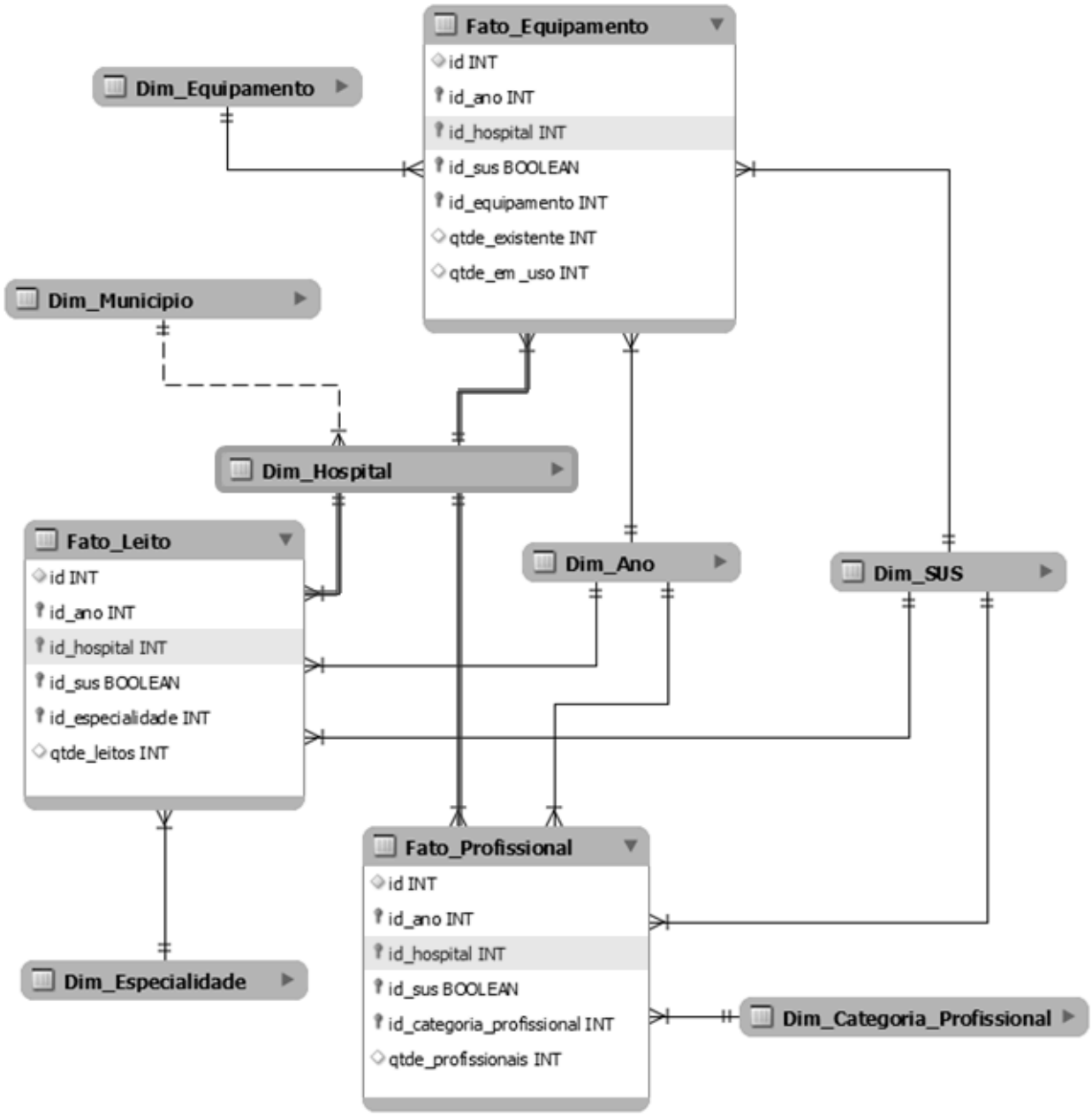


Figura 19 – DER da Constelação de Fatos por Hospital.

6.4. Criação do Processo ETL

O processo ETL - que consiste do conjunto de operações de extração, transformação e carregamento dos dados - foi implementado por meio do *software Pentaho Data Integration* (PDI). O PDI forneceu amplo amparo para a criação do processo ETL entre o BDint-GDMS e o DW-GDMS. Esta etapa do projeto foi realizada em paralelo à modelagem e criação das tabelas do DW-GDMS, ou seja, ao mesmo tempo em que os esquemas estrelas eram modelados e criados no DW-GDMS o processo ETL dos recém-criados esquemas eram implementados em conjunto.

O *software* PDI, mais especificamente o módulo *Spoon*, forneceu suporte ao processo ETL através de uma interface gráfica, permitindo implementar as etapas do processo de maneira prática. As principais funções do PDI são chamadas de **transformação** (*transformation*) e **tarefa** (*job*).

A **transformação** possui um conjunto de **passos** (*steps*) responsáveis por realizar o processamento dos dados, por exemplo, a obtenção dos dados de fontes legadas, a formatação e integração deles, e a carga para um repositório de dados de saída. Estes dados são transportados entre os **passsos** por meio de **saltos** (*hops*) que representam o fluxo dos dados desde a extração até a carga dos mesmos. O PDI também permite inserir **notas** (*note*) explicativas nas transformações. A Figura 20 apresenta um exemplo genérico de representação gráfica das etapas do processo ETL utilizando o módulo *Spoon* do PDI.

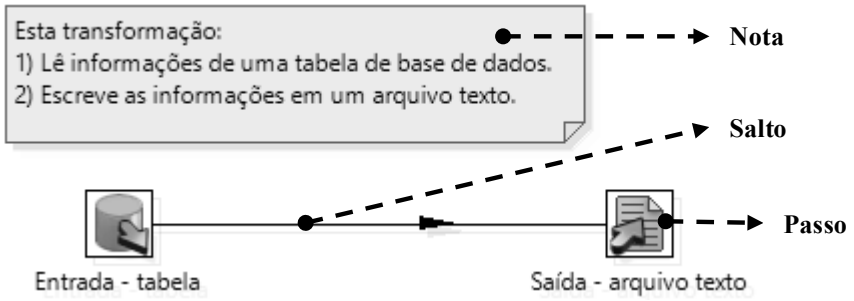


Figura 20 - Representação gráfica de uma transformação no PDI - *Spoon*.

A **tarefa** consiste de um conjunto de procedimentos para executar as **transformações**, dentre os quais incluem-se a iniciação (*start*) da **transformação** ou de outra **tarefa**, o agendamento delas (por exemplo, execução diária, semanal, mensal), o recebimento e envio de

mensagens de e-mail, configurações de comunicação em rede, gerenciamento de arquivos, dentre outras funções. A Figura 21 mostra um exemplo de representação gráfica de uma **tarefa** utilizando o módulo *Spoon* do *software* PDI.

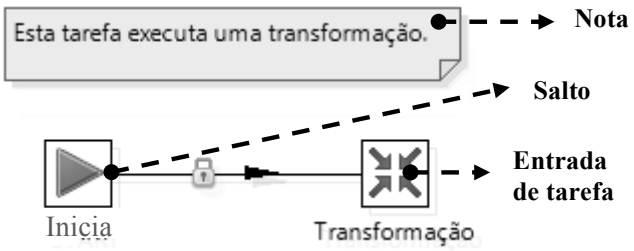


Figura 21 - Representação gráfica de uma **tarefa** no PDI - *Spoon*.

As subseções a seguir apresentam o processo ETL de dois esquemas estrelas do DW-GDMS, o primeiro é o processo ETL do esquema estrela de "Mortalidade" e o segundo é do esquema estrela de "Leito Hospitalar".

6.4.1. Carga do Esquema Estrela de Mortalidade

Primeiramente, como exemplo de ETL, o esquema estrela de "Mortalidade" possui cinco dimensões e uma tabela de fatos que foram carregadas com os dados oriundos do BDint-GDMS. Primeiro vamos explicar como foi realizado o processo ETL da "Dim_Municipio". A Figura 22 apresenta a transformação para extrair os dados de municípios do BD integrador e a carga na dimensão "Dim_Municipio" do DW-GDMS.

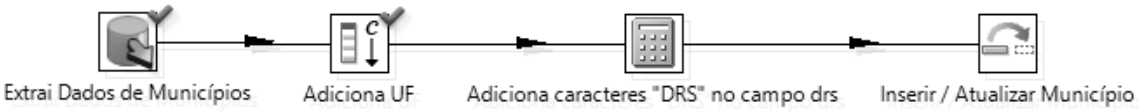


Figura 22 - Processo de carga dos dados da dimensão "Dim_Municipio".

O primeiro passo, através da linguagem SQL, realiza a extração dos dados de municípios das tabelas "Município", "DRS" e Regiao_Saude" do BDint-GDMS. A seguir, o segundo passo

insere em todas as linhas obtidas do primeiro passo os caracteres "SP" criando um campo chamado de "estado" (todos municípios são do Estado de São Paulo). Em seguida, o terceiro passo adiciona no início do campo "drs" os caracteres "DRS" com a finalidade de identificar todas as linhas deste campo como DRS e, deste modo, os valores contidos nesta coluna não confundem-se com os nomes de município, por exemplo, o nome do município de Ribeirão Preto confunde-se com o nome da DRS de Ribeirão Preto. Por fim, nesta transformação, o última passo realiza a carga de todos os dados na dimensão "Dim_Municipio" do DW-GDMS.

Avançando na criação do processo ETL, nós implementamos a transformação para a geração dos anos, biênios e décadas para a carga da dimensão "Dim_Ano". A Figura 23 mostra a transformação para gerar os dados para carga da dimensão "Dim_Ano".

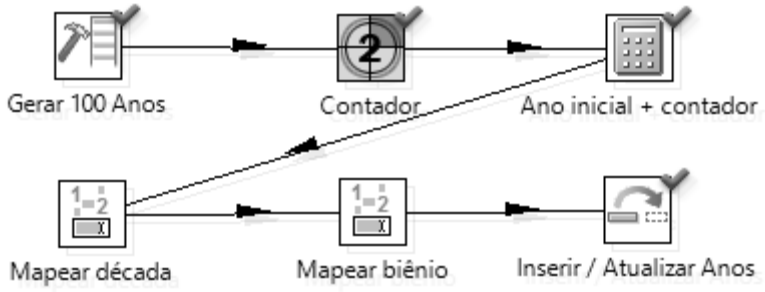


Figura 23 - Processo de geração dos dados da dimensão "Dim_Ano".

Como podemos visualizar na Figura 23, para a geração da carga foi necessário criar seis passos. O primeiro passo realiza a criação de 100 linhas e dois campos, um campo como identificador da tabela (id) e outro como ano inicial (ano_ini), os campos são criados com os valores iniciais zero e 2000 respectivamente. O segundo passo, denominado "Contador", realiza a contagem de uma sequência de zero a cem. Os valores gerados nesta sequência são transportados para o terceiro passo, onde os valores são somados às linhas do campo "id" e do campo "ano_ini", o resultado da soma deste último campo é atribuída a um novo campo, chamado "ano". Desta forma, estes primeiros passos realizam a contagem da sequência dos campos "ano" e "id", por exemplo, o "id" obtém como resultado os valores de suas linhas de zero a 99 e o ano de 2000 a 2099. Em seguida, no terceiro passo, os anos são mapeados para suas respectivas

décadas, similarmente ocorre no quarto passo, onde os anos são mapeados para seus respectivos biênios. Por fim, nesta transformação, os dados são carregados na dimensão “Dim_Ano”.

Seguindo na criação do processo ETL, nós realizamos a implementação da carga dos dados da dimensão “Dim_CID10”. O primeiro passo da transformação realiza a extração dos capítulos CID10 da tabela “CapCID10” do BDint-GDMS. Em virtude dos valores do campo descrição da tabela “CapCID10” conterem o número do capítulo juntamente com a descrição, nós realizamos a quebra deste campo com a finalidade de separá-lo em dois novos campos, capítulo e descrição. Por fim, nesta transformação, os capítulos e descrições do CID10 são carregados na dimensão “Dim_CID10”. A Figura 24 apresenta a transformação do processo ETL da dimensão “Dim_CID10”.

Figura 24 - Processo de carga dos dados da dimensão "Dim_CID10".

Seguindo no processo ETL, nós realizamos a criação da transformação para a carga dos dados da dimensão “Dim_Faixa_Etaria”. Esta transformação contém apenas dois passos para realizar a extração e carga dos dados. O primeiro realiza uma consulta em SQL no BDint-GDMS e seleciona os dados da tabela “Faixa_Etaria”, uma vez que os dados foram selecionados, eles são transportados para o último passo que é o de carregamento para a dimensão “Dim_Faixa_Etaria”. A Figura 25 apresenta a transformação que realiza a carga dos dados da dimensão “Dim_Faixa_Etaria”.

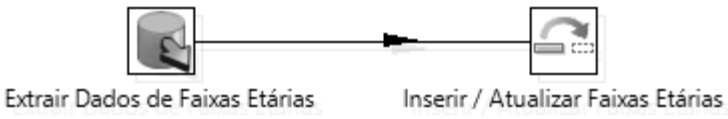


Figura 25 - Processo de carga dos dados da dimensão "Dim_Faixa_Etaria".

Mais adiante no processo ETL, nós criamos a transformação para a carga da dimensão “Dim_Sexo”. Nesta transformação, o primeiro passo realiza a extração dos dados de sexo de um arquivo CSV, estes dados são transportados para o segundo passo, onde é realizado o

carregamento na dimensão “Dim_Sexo”. A Figura 26 apresenta a transformação que realiza a carga dos dados da dimensão “Dim_Sexo”.

Figura 26 - Processo de carga dos dados da dimensão "Dim_Sexo".

Por fim, no processo ETL do esquema estrela de “Mortalidade”, nós implementamos a transformação para a carga da tabela de fatos “Fato_Mortalidade”. Nesta transformação, o primeiro passo extrai, por meio de uma consulta SQL, os campos de mortalidade da tabela “Morbimortalidade” do BDint-GDMS. O segundo passo filtra apenas os campos de interesse e ele seleciona as chaves para buscas das chaves primárias das dimensões de ano, município, faixa etária, sexo e capítulo CID10. Os cinco passos a seguir realizam a busca das chaves primárias das dimensões do esquema estrela para serem inseridas na tabela de fatos “Fato_Mortalidade”. Por fim, o último passo realiza a carga dos dados na tabela de fatos. A Figura 27 apresenta a transformação que realiza a carga dos dados da tabela de fatos “Fato_Mortalidade”.

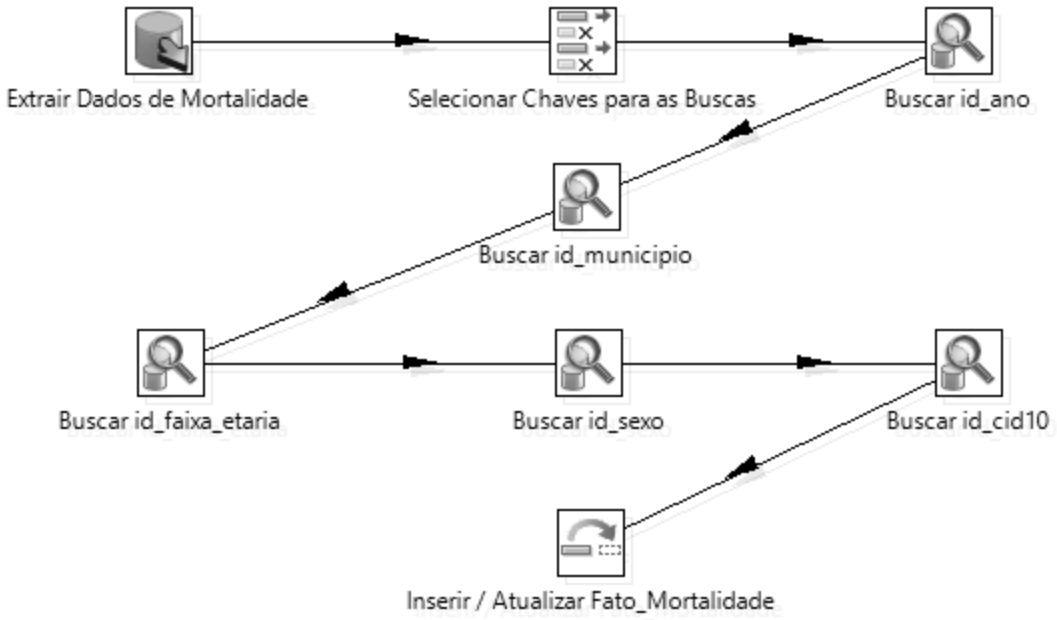


Figura 27 - Processo de carga dos dados da tabela de fatos "Fato_Mortalidade".

6.4.2. Carga do Esquema Estrela de Equipamento Hospitalar

O segundo exemplo de ETL é a carga dos dados do esquema estrela de “Equipamento Hospitalar”. Neste esquema nós realizamos a carga de três dimensões e uma tabela de fatos, o esquema também possui outras duas dimensões que são compartilhadas e já foram criadas transformações no esquema estrela anterior, que são as dimensões “Dim_Ano” e “Dim_Municipio”. Estas transformações se estendem para todos os esquemas estrelas que compartilham essas dimensões.

A primeira transformação que nós implementamos para este esquema foi a carga da dimensão “Dim_Hospital”. A Figura 28 apresenta a transformação do processo de carga da dimensão “Dim_Hospital”.

Figura 28 - Processo de carga dos dados da dimensão "Dim_Hospital".

Como podemos visualizar na Figura 28, o primeiro passo realiza a extração, por meio de uma consulta SQL, dos dados das tabelas “Hospital” e “Municipio” do BDint-GDMS, estes dados são transportados para o segundo passo, onde é realizada a busca das chaves do município para recuperar os respectivos municípios aos quais pertencem os hospitais. Por fim, os dados são carregados na dimensão “Dim_Hospital”.

A segunda transformação que nós implementamos para este esquema foi para a carga da dimensão “Dim_SUS”. Nesta transformação o primeiro passo realiza a extração dos dados de SUS de arquivo em formato XLS, a seguir os dados são transportados para o último passo onde ocorre a carga dos dados na dimensão. A Figura 29 apresenta a transformação que realiza a carga da dimensão “Dim_SUS”.

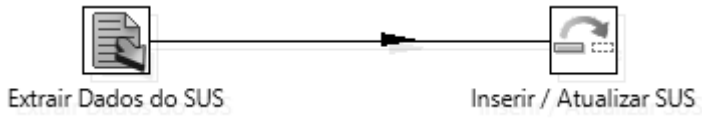


Figura 29 - Processo de carga dos dados da dimensão "Dim_SUS".

A terceira transformação que desenvolvemos para este esquema foi para a carga da dimensão “Dim_Equipamento”. O primeiro passo desta transformação realiza a extração dos dados, por meio de uma consulta SQL, da tabela “Equipamento” do BDint-GDMS. O segundo passo aplica uma formatação nos caracteres do campo “grupo”, esta formatação foi aplicada porque os caracteres estavam em maiúsculo, assim, este passo formata os caracteres para minúsculo mantendo apenas a primeira letra de cada palavra em maiúsculo. A Figura 30 apresenta a transformação que realiza a carga dos dados da dimensão “Dim_Equipamento”.

Figura 30 - Processo de carga dos dados da dimensão "Dim_Equipamento".

Por fim, neste exemplo, nós implementamos a transformação do processo de carga dos dados da tabela de fatos “Fato_Equipamento”. Seguindo a mesma lógica da primeira transformação para uma tabela de fatos, a transformação inicia extraindo os dados de uma tabela do BD integrador, neste caso o da tabela “Equipamento” depois de obtidos os dados eles são transportados para o segundo passo que realiza a seleção das chaves para as buscas das chaves nas dimensões, os três passos seguintes realizam as buscas das chaves nas dimensões “Dim_Ano”, “Dim_Equipamento” e “Dim_Hospital”. Depois de obtidas as chaves elas são carregadas na tabela de fatos “Fato_Equipamento” juntamente com as medidas. A Figura 31 apresenta a transformação que realiza o processo de carga dos dados da tabela de fatos “Fato_Equipamento”.

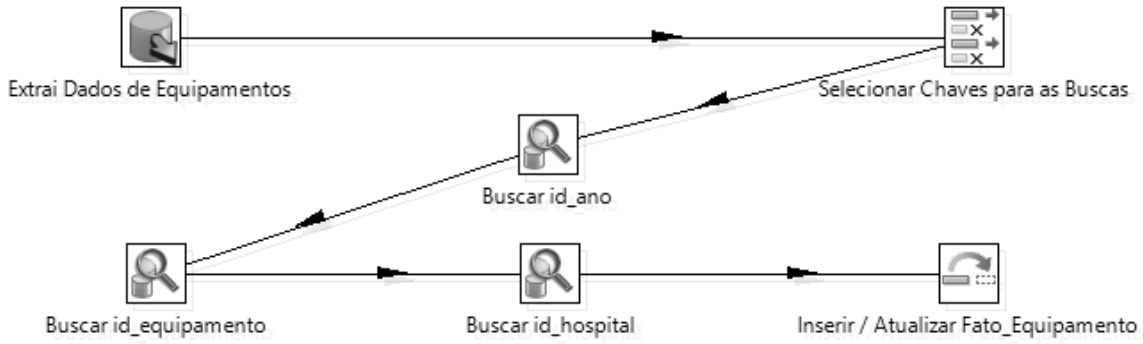


Figura 31 - Processo de carga dos dados da tabela de fatos "Fato_Equipamento".

6.5. Configuração do Servidor OLAP

A configuração do servidor OLAP foi desenvolvida por meio do software *Pentaho Schema Workbench* (PSW), o qual forneceu uma interface gráfica para projetar os cubos OLAP. O mecanismo do PSW processa requisições do tipo MDX (*multi-dimensional expressions* – expressões multidimensionais) (PENTAHO MONDRIAN DOCUMENTATION, 2015) que é uma linguagem interpretada pelo Mondrian utilizada para realizar consultas sobre os esquemas ROLAP (OLAP relacional). Esses esquemas são modelos de metadados em XML que são criados em uma estrutura especifica. Esses modelos em XML possuem uma estrutura do tipo cubo através da utilização das tabelas de fatos e dimensões encontradas no DW-GDMS, criado no PostgreSQL. A Figura 32 apresenta a interface exibindo um exemplo de esquema com os cubos criados para o DW-GDMS.

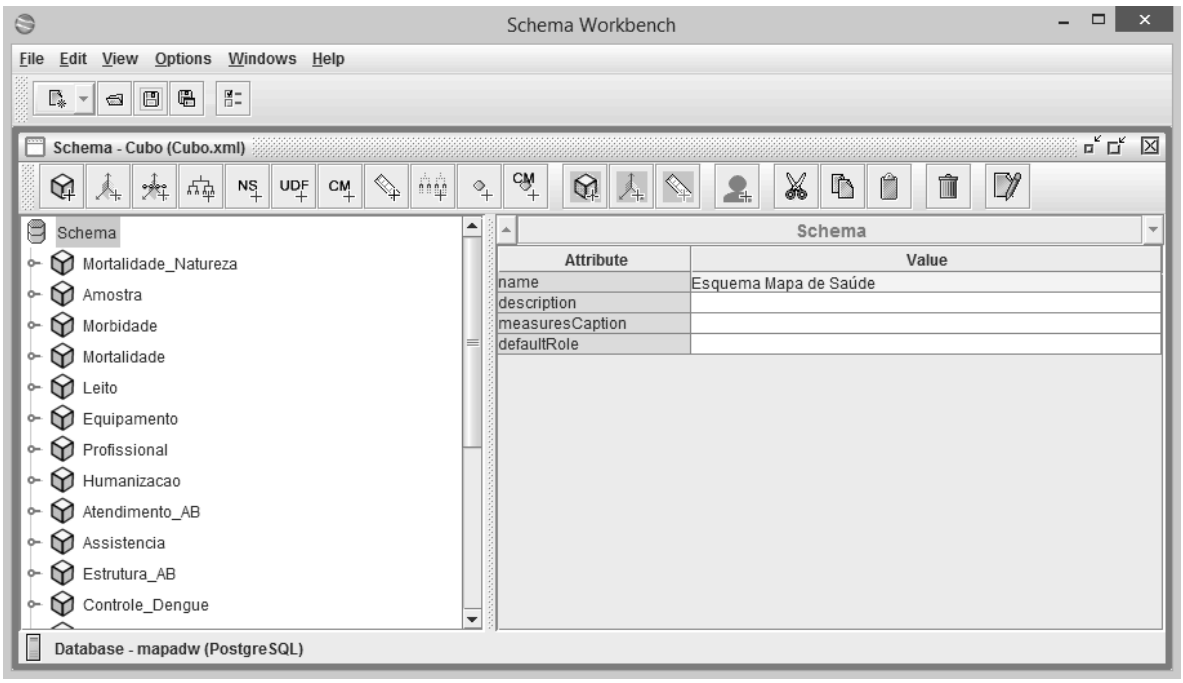


Figura 32 - Interface gráfica do *Pentaho Schema Workbench*.

Na Figura 32 podemos observar do lado esquerdo da janela os cubos que nós implementamos a partir dos esquemas estrelas do DW-GDMS. Cada cubo foi obtido a partir de um esquema estrela, por exemplo, um cubo para o esquema estrela "Mortalidade" e um para o esquema estrela "Equipamento" cujos DERs já foram apresentados na seção de modelagem do DW-GDMS.

A título de exemplo de como criamos os cubos de dados, podemos expandir o item do cubo de dados "Profissional" para visualizar detalhadamente a sua estrutura. A Figura 33 apresenta o cubo "Profissional" expandido para visualização detalhada da sua estrutura.

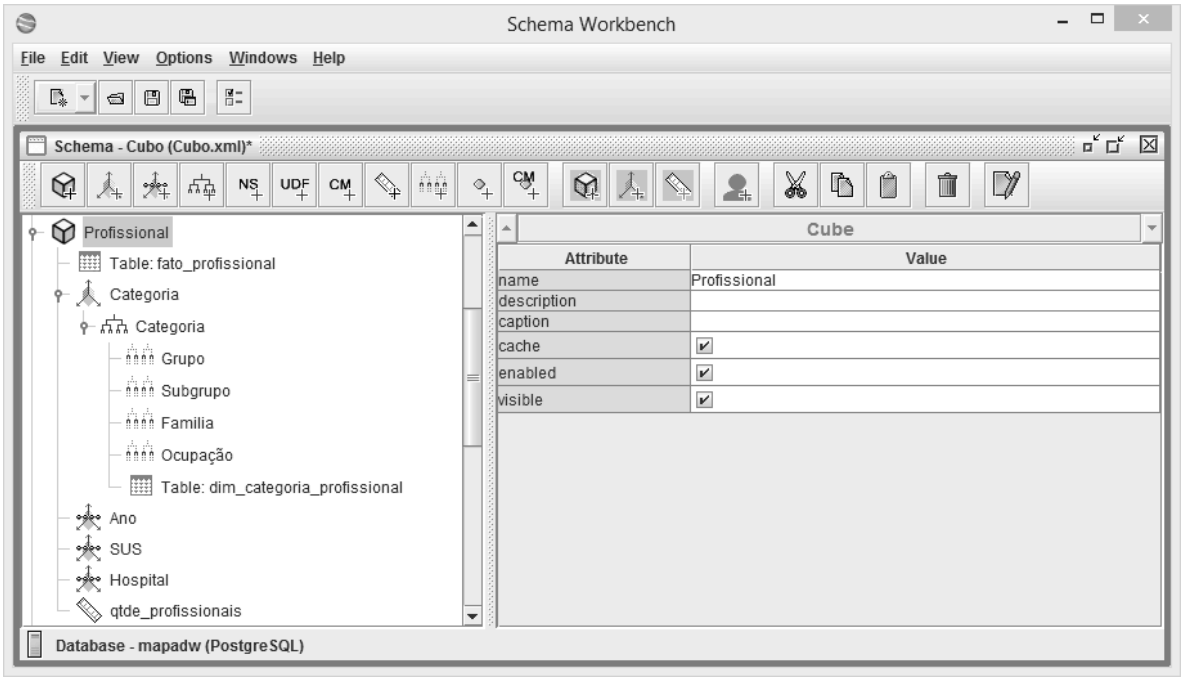


Figura 33 - PSW com o item do cubo de dados "Profissional" expandido.

Como se observa na Figura 33, o cubo "Profissional" contém um item referenciando a tabela de fatos "Fato_Profissional" criada no DW-GDMS e as dimensões Categoria, Ano, SUS e Hospital que se relacionam com esta tabela de fatos. As dimensões Ano, SUS e Hospital são compartilhadas, portanto nós criamos dimensões de uso geral para todos os cubos. Por outro lado, a dimensão Categoria é de uso específico deste cubo; assim, esta dimensão só é navegável por meio deste cubo. Adentrando na expansão do item da dimensão Categoria existem quatro níveis de hierarquia, a saber: ocupação; família; subgrupo; e grupo. O maior grão desta hierarquia é o nível "ocupação" e o menor grão é o nível "grupo".

Os cubos configurados geram modelos de metadados em XML. Cada esquema de cubos gera um arquivo XML que é interpretado pelo servidor OLAP juntamente com a base de dados do DW-GDMS. A Figura 34 apresenta uma parte do arquivo XML do esquema de cubos a qual representa o cubo "Profissional".

```
<Cube name="Profissional" visible="true" cache="true" enabled="true">
    <Table name="fato_profissional" schema="public"></Table>
    <Dimension type="StandardDimension" visible="true" foreignKey="id_cbo"
    highCardinality="false" name="Categoria">
        <Hierarchy name="Categoria " visible="true" hasAll="true" allMemberName="Todas Categorias"
        allMemberCaption="Todas Categorias" primaryKey="id_cbo" caption="Categoria">
            <Table name="dim_categoria_profissional" schema="public"></Table>
            <Level name="Grupo" visible="true" table="dim_categoria_profissional" column="grupo"
            ordinalColumn="grupo" type="String" uniqueMembers="false" levelType="Regular"
            hideMemberIf="Never"></Level>
            <Level name="Subgrupo" visible="true" table="dim_categoria_profissional" column="subgrupo"
            type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"></Level>
            <Level name="Familia" visible="true" table="dim_categoria_profissional" column="familia"
            type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"></Level>
            <Level name="Ocupa&#231;&#227;o" visible="true" table="dim_categoria_profissional"
            column="ocupacao" type="String" uniqueMembers="false" levelType="Regular"
            hideMemberIf="Never"></Level>
        </Hierarchy>
    </Dimension>
    <DimensionUsage source="Ano" name="Ano" visible="true" foreignKey="id_ano"
    highCardinality="false"></DimensionUsage>
    <DimensionUsage source="SUS" name="SUS" visible="true" foreignKey="id_sus"
    highCardinality="false"></DimensionUsage>
    <DimensionUsage source="Hospital" name="Hospital" visible="true" foreignKey="id_hospital"
    highCardinality="false"></DimensionUsage>
    <Measure name="qtde_profissionais" column="qtde_profissionais" datatype="Integer"
    aggregator="sum" caption="Profissionais" visible="true"></Measure>
</Cube>
```

Figura 34 - Cubo "Profissional" em XML.

6.6. Exemplos de Análises

Depois de desenvolvidos todos os esquemas estrelas e seus respectivos cubos, os modelos de metadados criados em XML são publicados no servidor *Pentaho Business Analytics* (PBA), o qual contém uma biblioteca *Web* desenvolvida em Java com um conjunto de aplicativos para criação de relatórios e execução de consultas analíticas sobre os dados do DW-GDMS. Dentre essas ferramentas, o JPivot foi o mais utilizado para realizar análises dinâmicas dos dados. O JPivot é desenvolvido em JSP (*JavaServer Pages*) e fornece uma tabela OLAP a qual permite que usuários realizem navegações OLAP, tais como: *slice-and-dice*, *drill-down*, e *roll-up*, etc. O apêndice A apresenta uma breve explicação de como funciona o servidor PBA e como pode ser utilizado o aplicativo JPivot para realizar algumas operações OLAP.

Como exemplos de consultas por meio da ferramenta desenvolvida, realizamos três análises baseadas em necessidades do dia-a-dia dos gestores de saúde.

Análise 1: Mostrar a quantidade de profissionais por estabelecimentos nos anos 2012 e 2013, sendo os estabelecimentos organizados de acordo com a hierarquia de seus níveis ("SP → RRAS13 → DRS Ribeirão Preto → Aquífero do Guarani → Municípios"). Dentro desta hierarquia, mais especificamente no nível de Região de Saúde, foram selecionas mais de uma região (Vale das Cachoeiras, Horizonte Verde e Aquífero do Guarani), isto feito também para municípios (Serrana, Serra Azul, São Simão, etc.). Além disso, nós aplicamos dois filtros, o primeiro condicionando a consulta a exibir somente profissionais da família "médicos clínicos" e o segundo selecionando somente os profissionais que trabalham no SUS. A Figura 35 apresenta a análise construída sobre o cubo "Profissional".

	Measures	
	Qtde Profissionais	
	Ano	
Estabelecimento	●2012	●2013
RRAS13	7.942	7.962
DRS RIBEIRÃO PRETO	3.414	3.430
VALE DAS CACHOEIRAS	380	380
HORIZONTE VERDE	733	749
AQUÍFERO GUARANI	2.301	2.301
Serrana	73	72
Serra Azul	16	11
São Simão	32	31
Santa Rosa de Viterbo	57	61
Santa Rita do Passa Quatro	96	110
Ribeirão Preto	1.914	1.903
Luís Antônio	20	20
Jardinópolis	37	37
Guatapará	14	13
Cravinhos	42	43

Slicer: [Familia=Médicos Clínicos] [SUS=SUS]

Figura 35 – Análise construída do cubo "Profissional".

Análise 2: Mostrar a análise das medidas de óbitos por local de ocorrência e por residência para os departamentos regionais de saúde, no ano de 2010 e para o capítulo XI da CID10, de "Doenças

do aparelho digestivo". Além disso, configurar e mostrar um gráfico de barras para as linhas da tabela desta análise.

Primeiramente, o processo de construção desta análise inicia-se com a escolha da "Análise de Mortalidade por Capítulos CID10" que está contida na pasta "II. Perfil de Morbimortalidade" e na subpasta "Mortalidades por Capítulos CID10" no servidor PBA. A Figura 36 apresenta o início da "Análise de Mortalidade por Capítulos CID10" sendo executada por meio do servidor PBA.

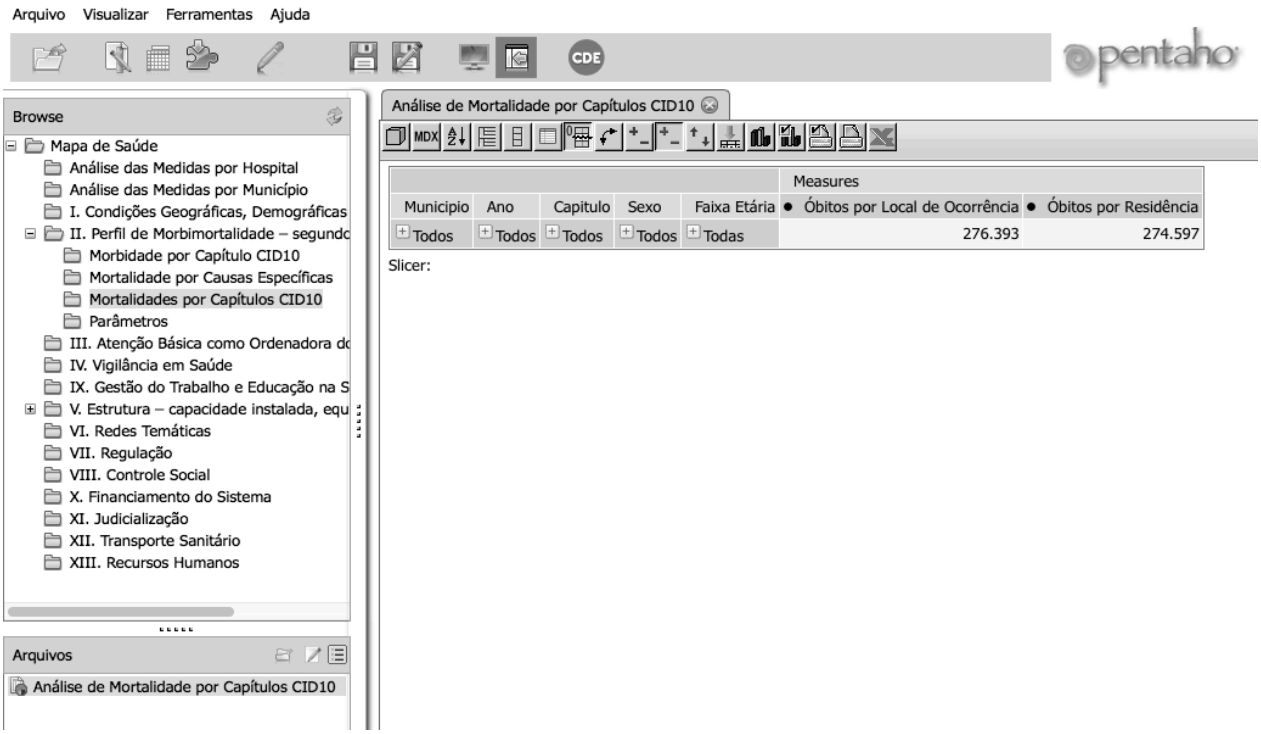


Figura 36 – Início da "Análise de Mortalidade por Capítulos CID10".

Depois de escolhida a análise, uma tabela dinâmica é apresentada por meio do aplicativo JPivot no servidor PBA. A tabela mostra, inicialmente, as dimensões (colunas) envolvidas nesta análise (por exemplo: Município, Ano, Capítulo, Sexo, Faixa Etária) e a soma total de suas agregações para as medidas de "Óbitos por Local de Ocorrência" e "Óbitos por Residência".

O segundo passo do processo de construção da análise é aplicar os filtros sobre as dimensões, neste caso selecionando o ano "2010" para a dimensão "Ano" e o capítulo "XI.

Doenças do aparelho digestivo” para a dimensão “Capítulo”. A Figura 37 apresenta a aplicação do filtro nas dimensões Ano e Capítulo da “Análise de Mortalidade por Capítulos CID10”. Para as dimensões de “Faixa Etária” e “Sexo” também foram determinados filtros, entretanto, para mostrar seus valores totais (por exemplo “todos os sexos” e “todas as faixas etárias”).

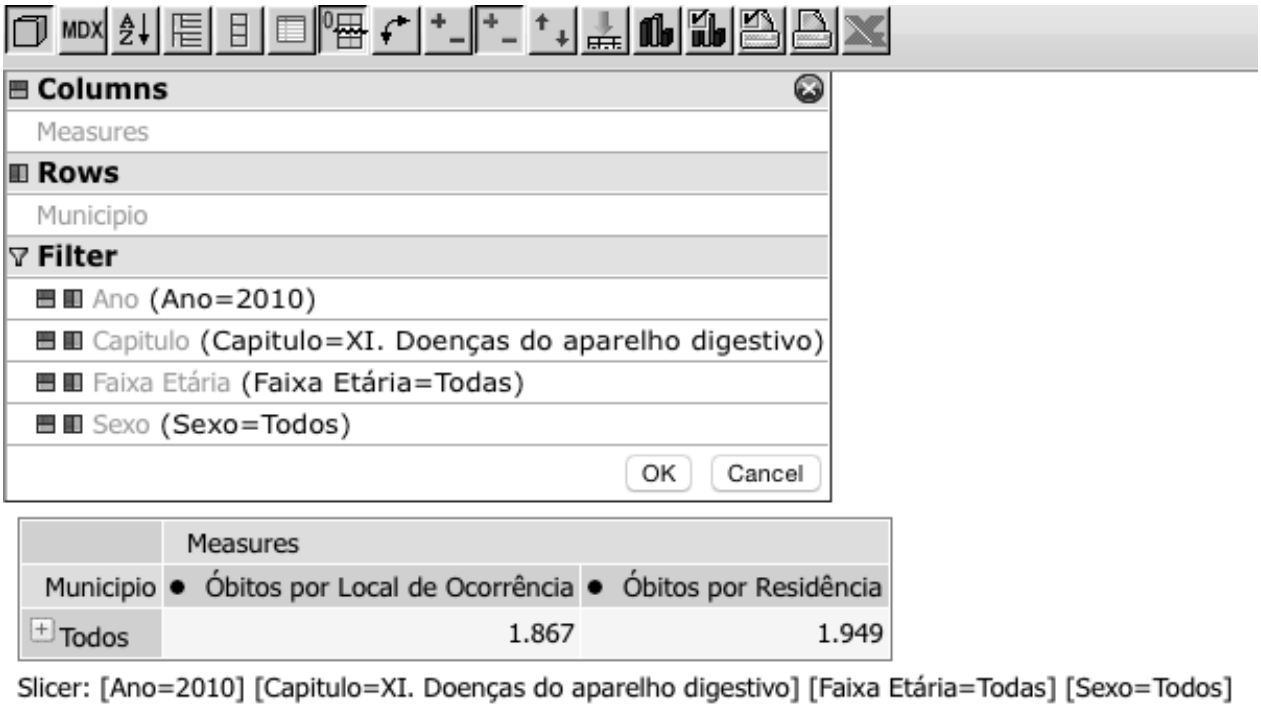
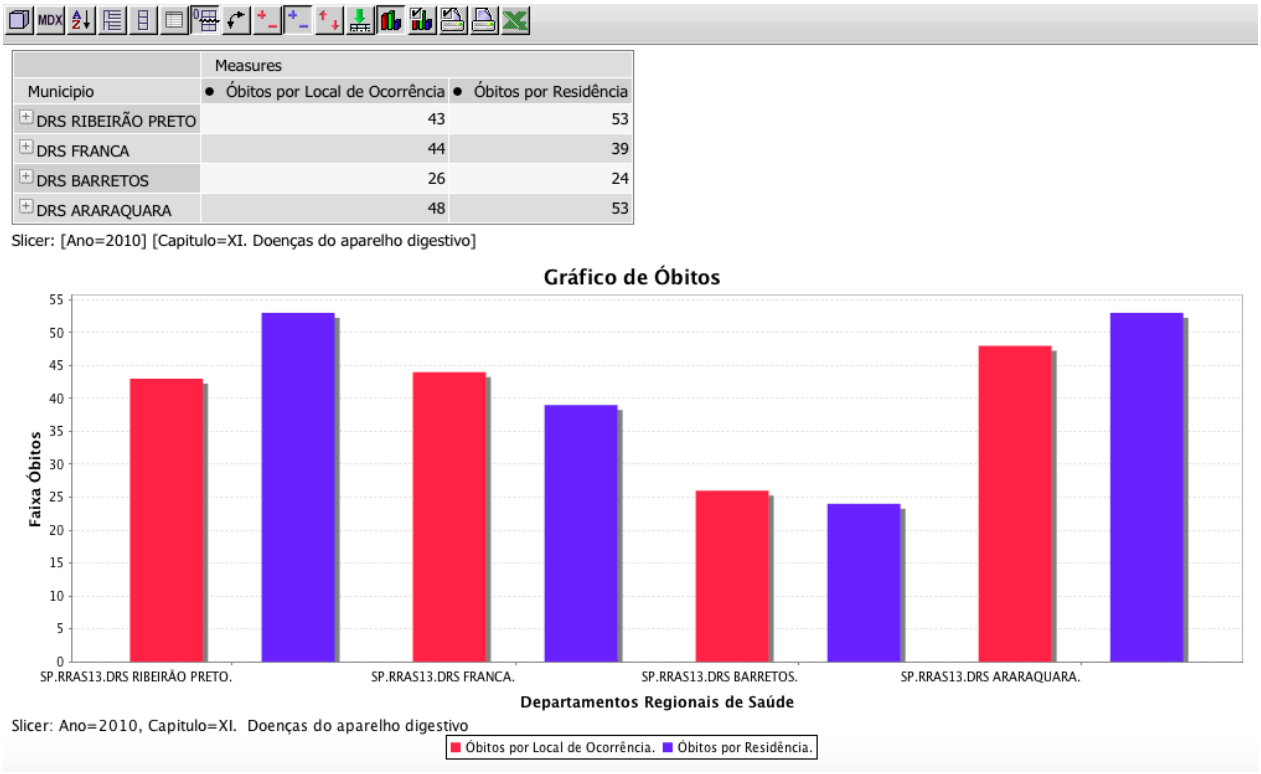


Figura 37 – Aplicação de filtros na “Análise de Mortalidade por Capítulos CID10”.

Por fim, o último passo do processo de construção desta análise consiste em detalhar a dimensão município para o nível de DRS, ou seja, utiliza-se a operação de *drill-replace* (perfurar substituir), a qual navega na hierarquia da dimensão detalhando seus níveis. Neste caso, a operação *drill-replace* é aplicada da seguinte forma: “Todos → SP → RRAS13”. Quando a operação é aplicada no nível “RRAS13” todas as DRSs desta RRAS são mostradas. Através do resultado desta análise podemos configurar e construir um gráfico. A Figura 38 apresenta o resultado final das operações aplicadas na “Análise de Mortalidade por Capítulos CID10” e seu respectivo gráfico.



	Measures	
Municipio	Óbitos por Local de Ocorrência	Óbitos por Residência
DRS RIBEIRÃO PRETO	43	53
DRS FRANCA	44	39
DRS BARRETOS	26	24
DRS ARARAQUARA	48	53

Slicer: [Ano=2010] [Capitulo=XI. Doenças do aparelho digestivo]

Figura 38 – Resultado da "Análise de Mortalidade por Capítulos CID10" e seu respectivo gráfico.

Análise 3: Mostrar algumas medidas que compartilham a dimensão município, tais como as medidas de amostras de água realizadas e satisfatórias, medidas de óbitos por residência e por local de ocorrência e medidas de internações por residência e por local de internação. A dimensão ano foi expandida na linha do Município de Ribeirão Preto, de modo que podemos observar as medidas por todos os anos, por décadas e por biênios através deste município. A Figura 39 apresenta a análise construída sobre o cubo da constelação de fatos por municípios.

		Measures					
Municipio	Ano	Amostas Realizadas	Amostas Satisfatórias	Óbitos por Residência	Óbitos por Local de Ocorrência	Internações por Residência	Internações por Local
AQUÍFERO GUARANI	Todos	7.914	6.792	4.461	5.281	123.692	170.017
Cravinhos	Todos	807	693	156	72	4.336	555
Guatapará	Todos	458	418	44	23	1.140	
Jardinópolis	Todos	1.087	850	210	131	5.025	768
Luís Antônio	Todos	504	395	46	21	1.178	
Ribeirão Preto	Todos	1.421	1.283	3.200	4.482	87.231	151.732
	2001-2010	915	819	1.604	2.236	36.919	62.493
	2008-2009	597	522	1.168	1.588	17.146	28.205
	2010-2011	318	297	436	648	19.773	34.288
	2011-2020	506	464	1.596	2.246	50.312	89.239
	2010-2011	243	223	1.239	1.761	16.442	29.261
	2012-2013	263	241	357	485	33.870	59.978
Santa Rita do Passa Quatro	Todos	740	676	211	185	5.769	5.731
Santa Rosa de Viterbo	Todos	783	781	156	90	4.254	2.343
Serra Azul	Todos	485	485	59	30	1.764	
Serrana	Todos	901	709	253	168	9.609	6.846
São Simão	Todos	728	502	126	79	3.386	2.042
HORIZONTE VERDE	Todos	8.170	6.679	2.631	2.131	59.292	43.494
VALE DAS CACHOEIRAS	Todos	4.656	4.091	1.018	782	29.814	20.017

Figura 39 – Análise construída da constelação de fatos por município.

6.7. Validação

O processo de validação do ambiente consistiu da aplicação de um questionário de usabilidade do sistema a 10 gestores responsáveis nas tomadas de decisões sobre os dados de saúde pública nas esferas municipal e regional, dentre eles: 3 diretores na área de planejamento da saúde, 3 secretários da saúde, 1 médico da área de planejamento da saúde, 1 especialista em gestão e planejamento em saúde, 1 assistente técnico de planejamento de ações em saúde, e 1 coordenador técnico do centro de informação e informática da saúde. O questionário foi adaptado do CSUQ (*Computer System Usability Questionnaire* – Questionário de Usabilidade de Sistema de Computador) da IBM (LEWIS, 1995). O questionário adaptado contêm 12 questões, sendo que cada questão corresponde a uma asserção e uma classificação em uma escala de pontos variando de 1 (“Discordo Fortemente”) a 7 (“Concordo Fortemente”). As questões são as seguintes:

Q1: Foi simples utilizar este sistema.

Q2: Sou capaz de realizar o meu trabalho de forma eficaz usando este sistema.

Q3: Me senti confortável usando este sistema.

Q4: Foi fácil aprender a usar este sistema.

Q5: Acredito que meu trabalho será mais produtivo com o uso deste sistema.

Q6: Quando eu cometi algum erro no sistema consegui resolver rápido e facilmente.

Q7: É fácil encontrar as informações que eu preciso.

Q8: As informações visuais que o sistema oferece são fáceis de entender.

Q9: As informações que o sistema oferece são eficazes em me ajudar a completar as tarefas.

Q10: A organização da interface do sistema é clara.

Q11: A interface do sistema é agradável.

Q12: No geral, eu estou satisfeito com o sistema.

Antes da aplicação do questionário, foi realizada uma explicação sobre o propósito do ambiente, seguida de uma demonstração de como utilizar algumas operações básicas do JPivot e mostrado alguns exemplos destas operações na prática de modo que os gestores pudessem ter uma noção prévia, antes deles praticarem o uso da ferramenta OLAP. Após a utilização da ferramenta, o questionário foi aplicado de forma que pudéssemos coletar a opinião deles em relação ao uso do ferramental desenvolvido.

As respostas do questionário aplicado foram colocadas em uma planilha e calculadas as respectivas médias das pontuações. Por meio desta planilha nós construímos um gráfico das médias das pontuações sobre cada questão. A Figura 40 apresenta o gráfico de barras das pontuações da médias para as questões. A média geral obtida, considerando as 12 questões foi 6,07.

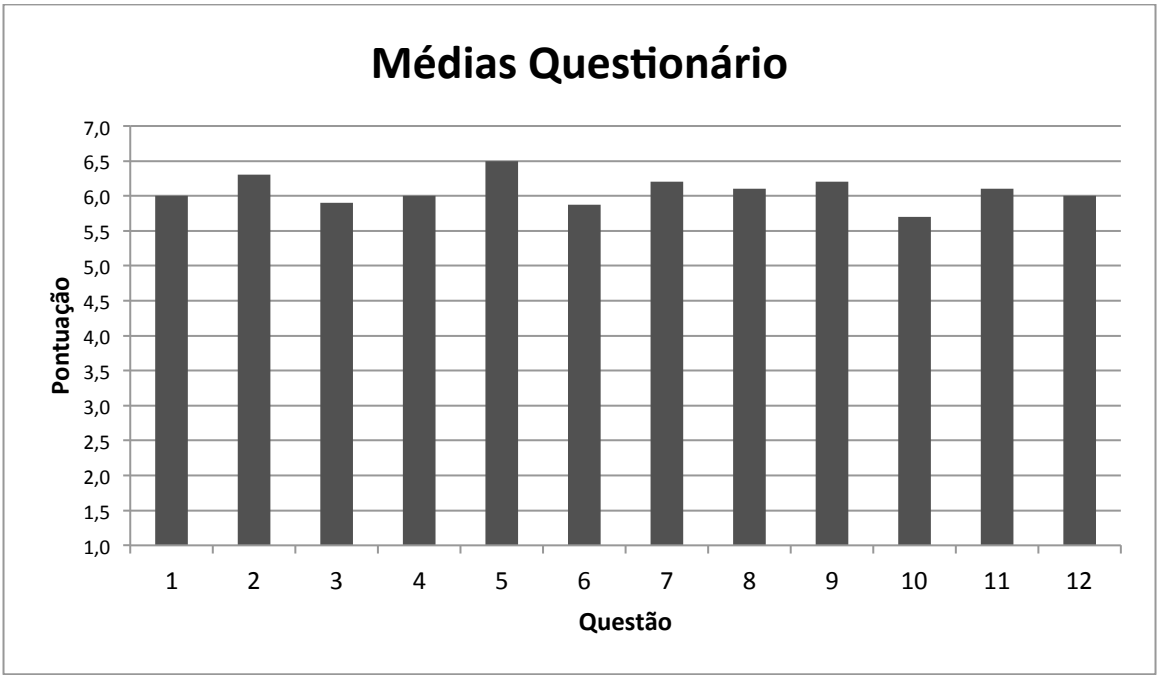


Figura 40 - Gráfico de barras das pontuações das médias para as questões.

7. Discussão

A implementação do ambiente de DWing com a finalidade de organizar, adotar um padrão e colaborar na análise dos dados para as tomadas de decisões em saúde pública mostrou-se desafiadora no que concerne o acesso às informações e à aceitação dos tomadores de decisões da saúde pública a nível regional.

Nas subseções a seguir, são discutidos alguns dos principais desafios, os resultados esperados e obtidos, e a reação dos analistas frente ao sistema construído.

7.1. Aquisição de fontes legadas

Podemos dizer que o Mapa de Saúde é uma iniciativa promissora em relação à organização e padronização dos dados de saúde pública porque oferece instruções e impõe uma conformidade para os dados. Em longo prazo, os documentos gerados pelo Mapa de Saúde podem ser utilizados para realizar análises de diversos tipos e por meio destas auxiliarem os técnicos, secretários, médicos e gestores, responsáveis por planejamento, a tomar decisões para melhor adequar os recursos da saúde pública. Todavia, o Mapa de Saúde, isoladamente, é incapaz de oferecer um meio de analisar os dados de maneira eficiente, dinâmica e flexível. Por este motivo a implementação do ambiente de DWing foi proposta. Este ambiente preenche esta lacuna que o Mapa de Saúde sozinho não pode preencher.

Entretanto, a construção deste ambiente analítico para atender essas necessidades exige o comprometimento das partes interessadas com diversas etapas para que o resultado seja bem sucedido. A primeira etapa do processo de desenvolvimento é a disponibilidade dos dados para a aquisição, de forma que eles possam ser extraídos, transformados de acordo com os requisitos do ambiente e enfim carregados para o repositório de DW. A primeira dificuldade encontrada, no início do projeto, foi em relação à aquisição dos dados de saúde pública. Apesar da maioria dos dados estarem bem especificados e descritos no Mapa de Saúde, a etapa de extração tornou-se complexa devido a fatores tais como: ausência de padrão, carência de referências em relação às fontes de dados e falta de permissão de acesso para algumas destas. Outro fator relevante foi a falta de orientação de como adquirir os dados. Algumas seções do Mapa de Saúde trazem referências às fontes de dados, contudo, outras não informam e isto gerou dificuldades que atrasaram o projeto.

No entanto, apesar de todas as dificuldades nós adotamos uma alternativa para resolver este problema de aquisição dos dados. A solução encontrada foi construir um mapeamento dos dados do Mapa de Saúde descrevendo em qual capítulo o dado se encontra, qual a fonte e o formato de origem, qual o grão e para qual tabela no BDint-GDMS este dado teria como destino. Deste modo, o mapeamento facilitou o processo de construção das transformações para a carga do DW-GDMS.

7.2. Avaliação de Ferramentas

O ambiente GDMS fornece uma variedade de funções e aplicações que podem colaborar com os gestores em consulta pré-definidas, em análises complexas e construção de relatórios textuais dos indicadores de saúde apresentados pelos documentos do Mapa de Saúde.

As ferramentas de acesso aos dados foram construídas com a finalidade de fornecer interfaces para todos os cubos OLAP criados no DW-GDMS. O servidor foi configurado de modo que pudéssemos aplicar análises testes sobre os dados carregados para o DW-GDMS e a verificação de alguns erros ou falhas no processamento da informação.

O servidor do Pentaho possui uma interface Web que permite navegar nos cubos OLAP, o JPivot forneceu recursos de acesso aos cubos multidimensionais, tais como as operações de detalhar e agregar os dados. Para acessar essa interface Web é necessária a autenticação com usuário e senha, assim, o servidor mantém os dados seguros restringindo o acesso somente a quem tem permissão.

Por fim, outro recurso importante para os usuários que utilizarão o sistema é a possibilidade de eles salvarem suas consultas quando acharem necessário. Isto possibilita que, a partir da consulta salva, eles possam realizar uma nova consulta sobre as mesmas visões já montadas anteriormente.

7.3. Validação

A validação do ferramental construído para o acesso aos dados do GDMS consistiu da aplicação de um questionário de usabilidade. O gráfico construído para as respostas das questões (Figura 40) mostra, de modo geral, que o ferramental foi bem aceito pelos gestores da área de saúde.

Primeiro, analisando o gráfico de pontuação média individualmente por questão, as duas questões que tiveram as melhores médias foram as questões 2 ("Sou capaz de realizar o meu trabalho de forma eficaz usando este sistema.") e 5 ("Acredito que meu trabalho será mais produtivo com o uso deste sistema."). Assim, podemos dizer que os usuários consideram a

ferramenta importante no sentido de melhorar seu trabalho em termos de eficácia e produtividade.

Além disso, nenhuma média ficou abaixo de 5,7 mostrando uma homogeneidade positiva nos quesitos de usabilidade. Por fim, a média geral das pontuações, considerando as 12 questões foi de 6,07 mostrando, portanto, que o GDMS foi muito bem aceito pelos usuários.

8. Conclusão

Por meio da realização e implementação deste trabalho podemos concluir que o arcabouço computacional proposto possibilita auxiliar os profissionais que estão envolvidos no planejamento e organização dos recursos, ações e serviços de saúde a tomar de decisões com base em informações de qualidade de forma rápido e fácil. Segundo os resultados apresentados a partir da avaliação realizada, podemos afirmar que o ferramental disponibilizado por este arcabouço teve uma forte aceitação por parte destes profissionais.

O ferramental desenvolvido permite a criação de análises sobre os dados apresentados pelo Mapa de Saúde de diferentes formas, sejam estas através de uma tabela dinâmica que se ajusta de acordo com a necessidade do profissional utilizando o ferramental ou por meio de gráficos que permitem realizar consultas e fazer análises de tendências das informações no tempo.

Além disso, o ferramental também possibilita que o usuário salve suas consultas no servidor Web, de forma que um conjunto customizado de consultas possa ser criado para diferentes categorias de usuários, objetivos e contextos. Ademais, considerando o acesso à ferramenta, podemos também dizer que as consultas construídas e as análises salvas estão seguras em relação à confidencialidade, isto porque o sistema possui um módulo de autenticação para manter seguras as análises dos profissionais que fazem uso do ferramental.

Nosso trabalho também apontou uma possível dificuldade na obtenção dos dados para realizar carga no sistema, isto porque as informações apresentadas pelo Mapa de Saúde estão

distribuídas de forma não padronizadas em diferentes formatos e em diversas fontes distintas da saúde. Portanto, nosso arcabouço serve como uma forma inicial de estabelecer uma organização sobre estas informações de saúde.

Podemos concluir que o objetivo da presente proposta, de contribuir com um ambiente para auxílio à análise e à tomada de decisão em saúde foi atingido de forma efetiva e satisfatória, no sentido de facilitar e flexibilizar a atividade de análise da informação, no processo de gestão em saúde.

9. Trabalhos Futuros

Os resultados obtidos atendem grande parte do arcabouço computacional GDMS definido para este projeto, com exceção apenas do componente sugerido para alimentação de dados por parte dos municípios. Assim, uma das necessidades de melhoria para este projeto diz respeito à inserção e atualização dos dados no BD integrador. O fornecimento de dados por parte dos municípios é imprescindível no que diz respeito à manutenção dos dados do GDMS. A manutenção dos dados corresponde à atualização e à inserção de novas informações no BD integrador do GDMS e, consequentemente, quando os dados são carregados para o GDMS–DW torna-se possível a construção de análises sobre dados mais atuais além dos dados históricos pré-existentes.

A ampliação da abrangência geográfica dos dados para outros Estados do Brasil é importante para a utilização da capacidade do GDMS em atender os objetivos do Mapa de Saúde, o qual foi projetado pelo Governo Federal em âmbito nacional. Entretanto, esta ampliação exige um grande esforço no que diz respeito à obtenção de dados por parte de municípios de outros Estados do Brasil.

Referências Bibliográficas

ALVES, D. Da Modelagem Matemática e Simulação Computacional a Construção de Sistemas de Informação em Saúde. Tese de livre-docência submetida à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, 2015.

BANEK, M.; TJOA, A. M.; STOLBA, N. Integrating different grain levels in a medical data warehouse federation. **Data Warehousing and Knowledge Discovery**. Springer Berlin Heidelberg, p. 185-194, 2006.

BONTEMPO, C.; ZAGELOW, G. The IBM data warehouse architecture. **Communications of the ACM**, New York, v. 41, n. 9, p. 38-48, 1998.

BORGES, F. Q. Gestão da Informação no Sistema Único de Saúde. **Revista de Administração FACES Journal**, Belo Horizonte, v. 13, n. 2, p. 83-98, 2014.

BOUMAN, R.; DONGEN, J. V. **Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL**. Wiley Publishing, 2009.

CABIBBO, L.; TORLONE, R. Querying multidimensional databases. **Database programming languages**. Springer Berlin Heidelberg, p. 319-335, jan. 1998.

CASTERS, M.; BOUMAN, R.; DONGEN, J.V. **Pentaho Kettle solutions: building open source ETL solutions with Pentaho Data Integration**. John Wiley & Sons. 2010.

CLASSIFICAÇÃO BRASILEIRA DE OCUPAÇÕES. Disponível em: < http://www.mtecbo.gov.br/cbosite/pages/home.jsf>. Acesso em: 01 mar. 2015.

CHAUDHURI, S.; DAYAL, U. An overview of data warehousing and OLAP technology. **ACM Sigmod record**, New York, v. 26, n. 1, p. 65-74, 1997.

CIFERRI, C.D.A. **Distribuição dos Dados em Ambientes de Data Warehousing: O Sistema WebD^{2}W e Algoritmos voltados à Fragmentação Horizontal dos Dados**. 2002. Tese (Doutorado) - CIn, Universidade Federal de Pernambuco, Recife, 2002.

EINBINDER, J. S.; SCULLY, K. W.; PATES, R. D.; SCHUBART, J. R.; REYNOLDS, R. E. Case study: a data warehouse for an academic medical center. **Journal of Healthcare Information Management**, San Francisco, v. 15, n.2, p. 165-176, 2001.

EVANS, R. S.; LLOYD, J. F.; PIERCE, L. A. Clinical Use of an Enterprise Data Warehouse. **AMIA Annual Symposium Proceedings**, Bethesda, v. 2012, p. 189-198, 2012.

FEUERWERKER, L. C. M.; CECÍLIO, L. C. O. O hospital e a formação em saúde: desafios atuais. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 12, n. 4, p. 965-971, 2007.

GIOIA, A.; CAZZIN, G.; DAMIANI, E. SpagoBI: a distinctive approach in open source business intelligence. **Digital Ecosystems and Technologies, 2008. DEST 2008. 2nd IEEE International Conference on. IEEE,** New York, 2008. p. 592-595.

GOLFARELLI, M. Open source BI platforms: a functional and architectural comparison. **Data Warehousing and Knowledge Discovery**. Springer Berlin Heidelberg, p. 287-297, 2009.

GOPALKRISHNAN, V.; LI, Q.; KARLAPALEM, K. Star/snow-flake schema driven object-relational data warehouse design and query processing strategies. **Data Warehousing and Knowledge Discovery**. Springer Berlin Heidelberg, p. 11-22, 1999.

GRAY, P.; WATSON, H. J. Present and future directions in data warehousing. **ACM SIGMIS Database**, New York, v. *29,* n. 3, p. 83-90, 1998.

GUPTA, H.; HARINARAYAN, V.; RAJARAMAN, A.; ULLMAN, J. D. Index selection for OLAP. **Data Engineering, 1997. Proceedings. 13th International Conference on**. IEEE, Birmingham, p. 208-219, 1997.

HYDE J. Mondrian Architeture. Disponivel em: <http://mondrian.pentaho.org/documentation/architecture.php>. Acesso em: 26 jan. 2014.

INMON, W. H. **Building the data warehouse**. Wiley. com, 2005.

INMON, W. H. Information architecture for the 90's: legacy systems, operational data store, data warehouse. **PRISM Tech Topic**, v. 1, n. 13, p. 118-124, 1993.

JASPERSOFT. Disponível em: <http://www.jaspersoft.com>. Acesso em: 29 jan. 2014.

KERKRI, E. M.; QUANTIN, C.; ALLAERT, F. A.; COTTIN, Y.; CHARVE, P.; JOUANOT, F.; YÉTONGNON, K. An approach for integrating heterogeneous information sources in a medical data warehouse. **Journal of Medical Systems**, v. 25, n. 3, p. 167-176, 2001.

KIMBALL, R.; CASERTA, J. **The data warehouse ETL toolkit**. Wiley Publishing, Inc, 2004.

KIMBALL, R.; ROSS, M. **The data warehouse toolkit: the complete guide to dimensional modelling**. John Willey & Sons, Inc, 2002.

LEWIS, J. R. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. **International Journal of Human-Computer Interaction**, v. 7, n. 1, p. 57-78, 1995.

LOPES, J. M. L. **Avaliação da qualidade de informação no registo da cirurgia segura.** 2014. Dissertação (Mestrado) – Engenharia Biomédica, Universidade do Minho, Portugal, 2014.

LUJÁN-MORA, S.; TRUJILLO, J.; SONG, I. Y. A UML profile for multidimensional modeling in data warehouses. **Data & Knowledge Engineering**, v. *59,* n. (3), p. 725-769, 2006.

OLIVA, S. Z. ; Miyoshi, N. S. B. ; Dias, T. F. F. ; ALVES, D. ; Felipe, J. C. Proposta de Framework de Data Warehousing para Integração de Dados da Saúde Pública. Medicina (Ribeirao Preto. Online), v. 47, p. 83-88, 2014.

OPENI. Disponível em: <http://openi.org>. Acesso em: 29 jan. 2014.

PALO. Disponível em: <http://www.palo.net>. Acesso em: 29 jan. 2014.

PENTAHO. Disponível em: <http://www.pentaho.com>. Acesso em: 29 jan. 2014.

PENTAHO MONDRIAN DOCUMENTATION. Disponível em: < http://mondrian.pentaho.com/documentation/mdx.php>. Acesso em: 26 abr. 2015.

PEREIRA, E. **Apoio à decisão em ginecologia e obstetrícia.** 2014. Dissertação (Mestrado) – Engenharia Biomédica, Universidade do Minho, Portugal, 2014.

POE, V.; BROBST, S.; KLAUER, P. Building a data warehouse for decision support. Prentice-Hall, Inc., 1997.

PORTAL DA LEGISLAÇÃO GOVERNO FEDERAL. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/decreto/D7508.htm>. Acesso em: 30 jun. 2014.

POSTGRESQL GLOBAL DEVELOPMENT GROUP. PostgreSQL. Disponível em: <http://www.postgresql.org>. Acesso em: 26 jan. 2014.

PRAT, N.; AKOKA, J.; COMYN-WATTIAU, I. A UML-based data warehouse design method. **Decision support systems**, v. 42, n. 3, p. 1449-1473, 2006.

SAMOS, J.; SALTOR, F.; SISTAC, J.; BARDES, A. Database architecture for data warehousing: an evolutionary approach. **Database and Expert Systems Applications***,* Springer Berlin Heidelberg, p. 746-756, jan. 1998.

SCHERER, M. D. A.; PIRES, D.; SCHWARTZ, Y. Trabalho coletivo: um desafio para a gestão em saúde. **Revista Saúde Pública**, v. 43, n. 4, p. 721-725, 2009.

SECRETARIA DE ESTADO DA SAÚDE. SÃO PAULO. Disponível em: <http://www.saude.sp.gov.br/ses/perfil/gestor/documentos-de-planejamento-em-saude/mapa-de-saude>. Acesso em: 6 jan. 2014.

SOUZA, R. C; FREIRE, S. M. Integração de dados ambulatoriais de quimioterapia e radioterapia registrados nas bases de dados do SUS. Congresso Brasileiro de Engenharia Biomédica – CBEB. Uberlândia, p. 1904-1907, out. 2014.

SPAGOBI. Disponível em: <http://www.spagoworld.org/xwiki/bin/view/SpagoBI>. Acesso em: 29 jan. 2014.

TERESO, M.; BERNARDINO, J. Open source business intelligence tools for SMEs. **Information Systems and Technologies (CISTI), 2011 6th Iberian Conference on**. **IEEE**, p. 1-4, 2011.

THOMSEN, C.; PEDERSEN, T. B.. A survey of open source tools for business intelligence. **Data Warehousing and Knowledge Discovery**. Springer Berlin Heidelberg, p. 74-84, 2005.

VADUVA, A.; DITTRICH, K. R. Metadata management for data warehousing: between vision and reality. In: **Database Engineering & Applications, 2001 International Symposium on**. IEEE, p. 129-135, 2001.

VANILLA PLATFORM. Disponível em: <http://www.bpm-conseil.com/content/vanilla-platform>. Acesso em: 29 jan. 2014.

WU, M. C.; BUCHMANN, A. P. Research Issues in Data Warehousing. **Proceedings of The German Database Conference**, Ulm, Germany, p. 61-82, mar. 1997.

ZENG, L.; XU, L.; SHI, Z.; WANG, M.; WU, W. Techniques, Process, and Enterprise Solutions of Business Intelligence. **IEEE International Conference on Systems, Man and Cybernetics**, p. 4722-4726, 2006.

APÊNDICE A - Operações OLAP Utilizando o JPivot do Servidor PBA

Com a finalidade de explicar como funciona o servidor PBA, nós selecionamos o cubo de dados "Mortalidade_Natureza" do DW para mostrar como é a interface inicial da ferramenta, e também para apresentar as principais operações OLAP que podemos realizar por meio do aplicativo JPivot. A Figura 41 apresenta a interface inicial do aplicativo Web do servidor PBA.

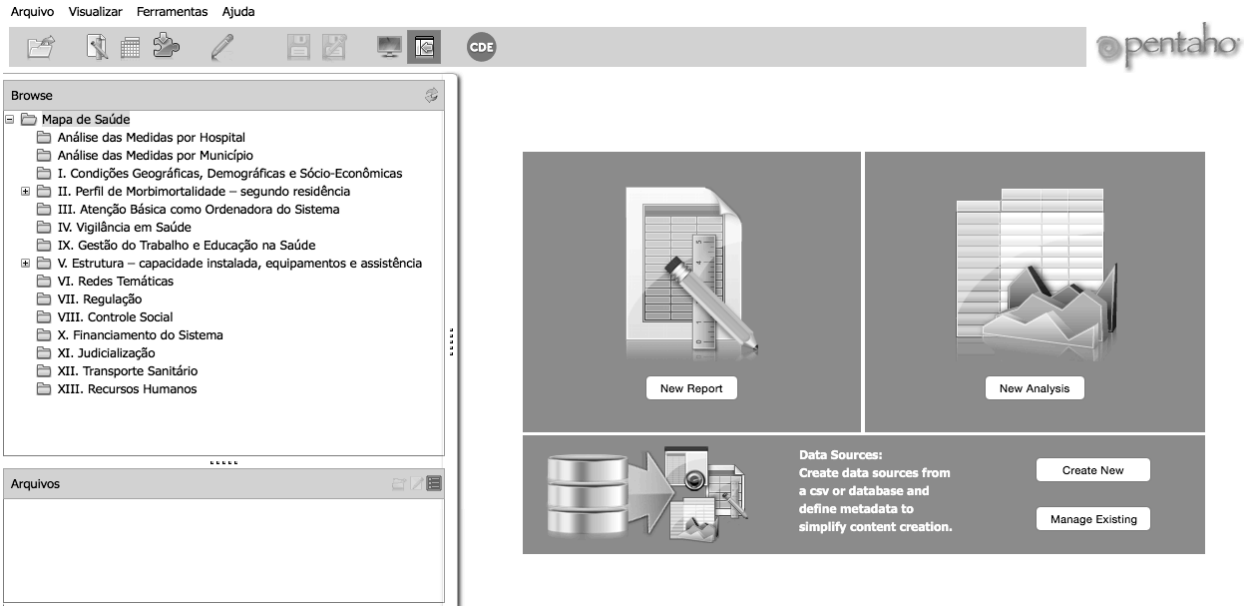


Figura 41 - Interface inicial do aplicativo Web do servidor PBA.

Na Figura 41, podemos observar a interface inicial do PBA, que contém uma janela do lado esquerdo que mostra pastas criadas no servidor e uma janela do lado direito que mostra opções de criar novos relatórios e novas análises, bem como criar novas fontes de dados e gerenciar as conexões com fontes de dados existentes. As pastas criadas no servidor seguem de modo similar a estrutura de capítulos definidos no Mapa de Saúde, com exceção das duas primeiras pastas que são análises baseadas nas constelações de fatos por Hospital e Município.

A partir da interface inicial do PBA podemos selecionar algumas das pastas para visualização das análises dos cubos. Por exemplo, selecionando a pasta do capítulo II, de perfil de

morbimortalidade, e a subpasta de mortalidade por causas específicas encontramos a análise do cubo “Mortalidade_Natureza”. A Figura 42 mostra a análise do cubo “Mortalidade_Natureza” executada por meio do JPivot no servidor PBA.

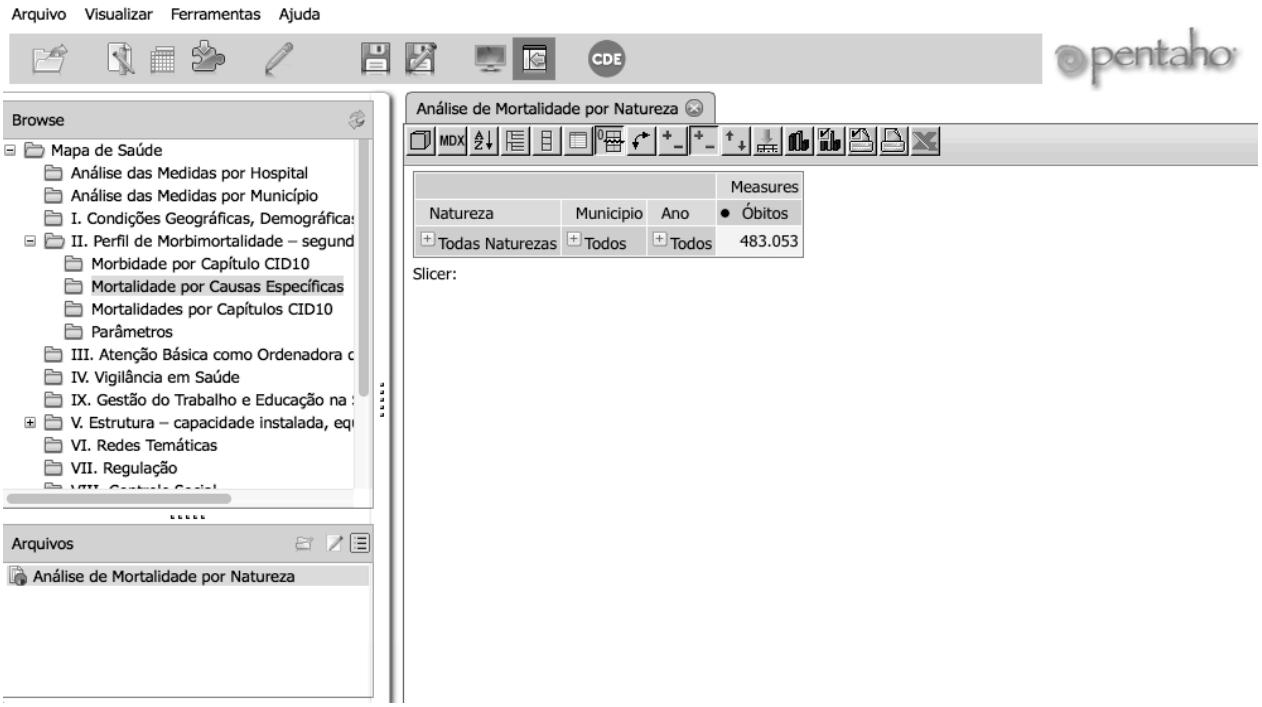


Figura 42 – Análise do cubo “Mortalidade_Natureza”.

Podemos realizar outras análises sobre este mesmo cubo de forma rápida e dinâmica, dentre elas navegarmos nas dimensões por meio de algumas operações oferecidas pelo JPivot, tais como: *drill-member* (perfurar membros), *drill-position* (perfurar posição) , *drill-replace* (perfurar repor) *drill-through* (perfurar através), *swap-axes* (trocar eixos). Estas operações são equivalentes às operações apresentadas na teoria das operações OLAP. O *drill-member*, o *drill-position* e o *drill-replace* são similares às operações *drill down* e *drill up* porque elas realizam tanto a função de detalhar os dados (*drill down*) quanto de agregá-los (*drill up*), entretanto, com algumas particularidades. O *drill-member* permite navegar nos níveis de agregação dos dados por meio da expansão e encolhimento dos níveis das colunas (dimensões) da tabela, porém, depois de expandido o nível de uma das colunas se em seguida outra coluna tiver um de seus níveis

expandido todos os níveis daquela coluna recém-selecionada serão expandidos. A Figura 43 apresenta um exemplo da operação *drill-member*.

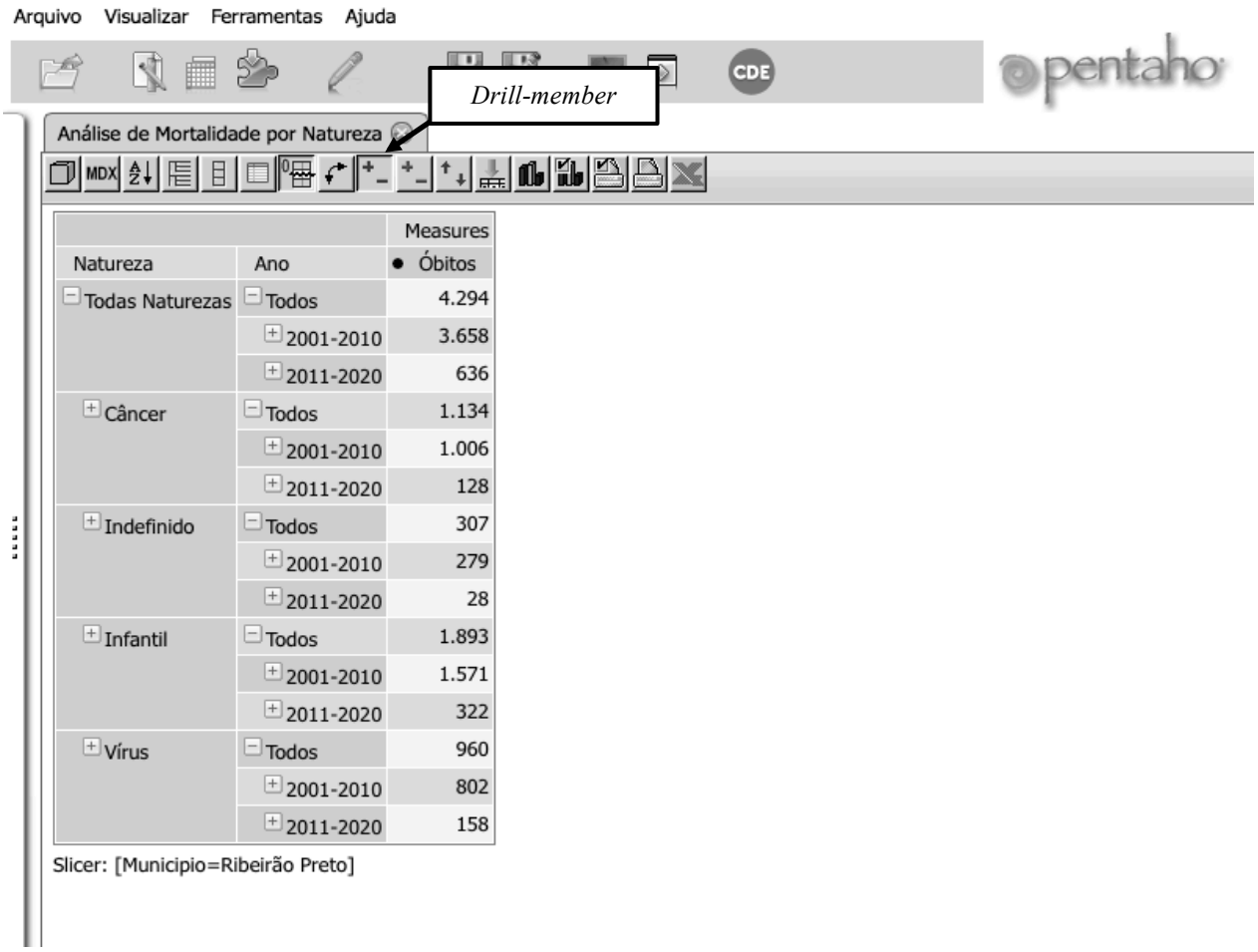


Figura 43 – Exemplo da operação *drill-member*.

Na Figura 43, como podemos observar, o botão *drill-member* foi clicado para aplicar a operação sobre a tabela de análise do cubo, depois de clicado o botão a operação foi ativada. Posteriormente, o nível "Todas" (o nível mais agregado) na coluna "Natureza" foi expandido tornando visíveis os seus subníveis. Após isso, o nível "Todos" na coluna "Ano" foi expandido e, consequentemente, todas as linhas da coluna "Ano" exibiram seus subníveis. Diferentemente ocorre na operação *drill-position*, que neste mesmo caso quando expandido o nível "Todos" na coluna "Ano" ao invés de todas as linhas exibirem seus subníveis, apenas o nível, ao qual foi

clicado, que exibiria seus subníveis. A Figura 44 apresenta a operação *drill-position* aplicada ao mesmo caso mostrado na operação *drill-member*.

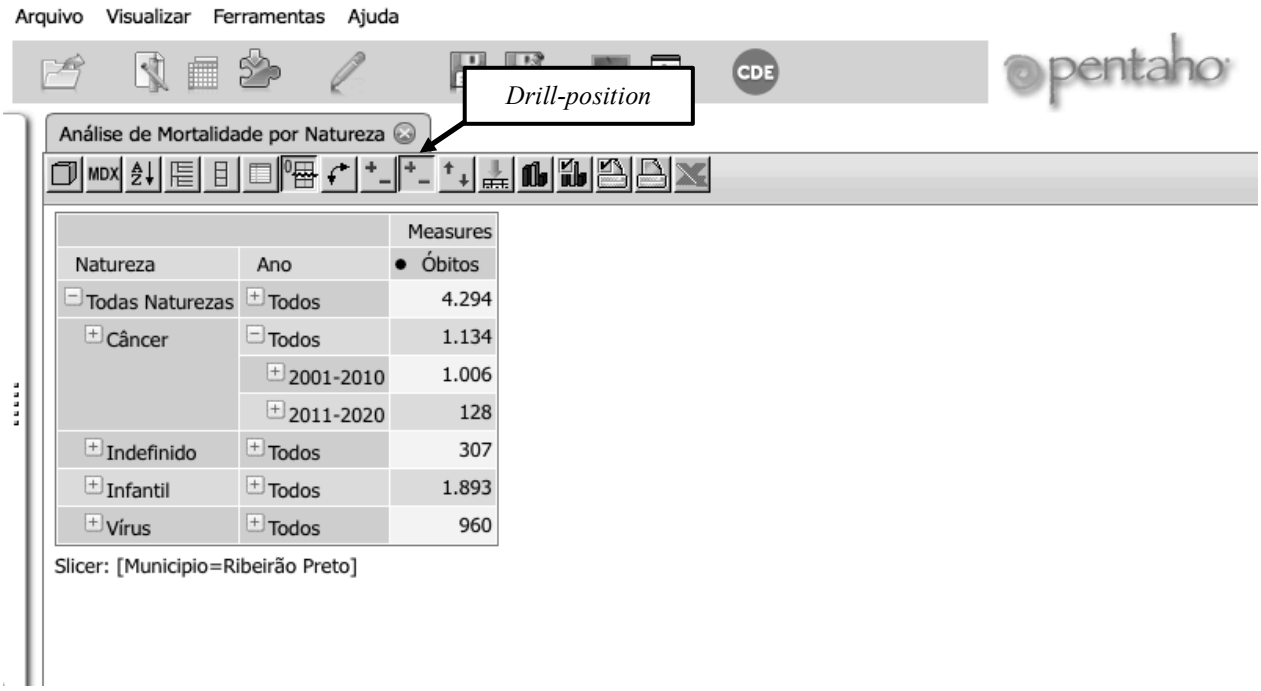


Figura 44 – Exemplo da operação *drill-position*.

Como podemos visualizar na Figura 44, a mesma tabela e seleção inicial da operação *drill-member* foram utilizadas, porém, como diferencial da consulta da operação *drill-position* quando o nível "Todos" da coluna "Ano" foi expandido apenas os subníveis do nível clicado que foram exibidos.

Outra operação que podemos aplicar na análise do cubo profissional é a operação *drill-replace* que se assemelha com a operação *drill-position* diferindo apenas que na *drill-replace* quando clicamos para expandir um nível, este é substituído por seus subníveis. A Figura 45 apresenta um exemplo da operação *drill-replace*.

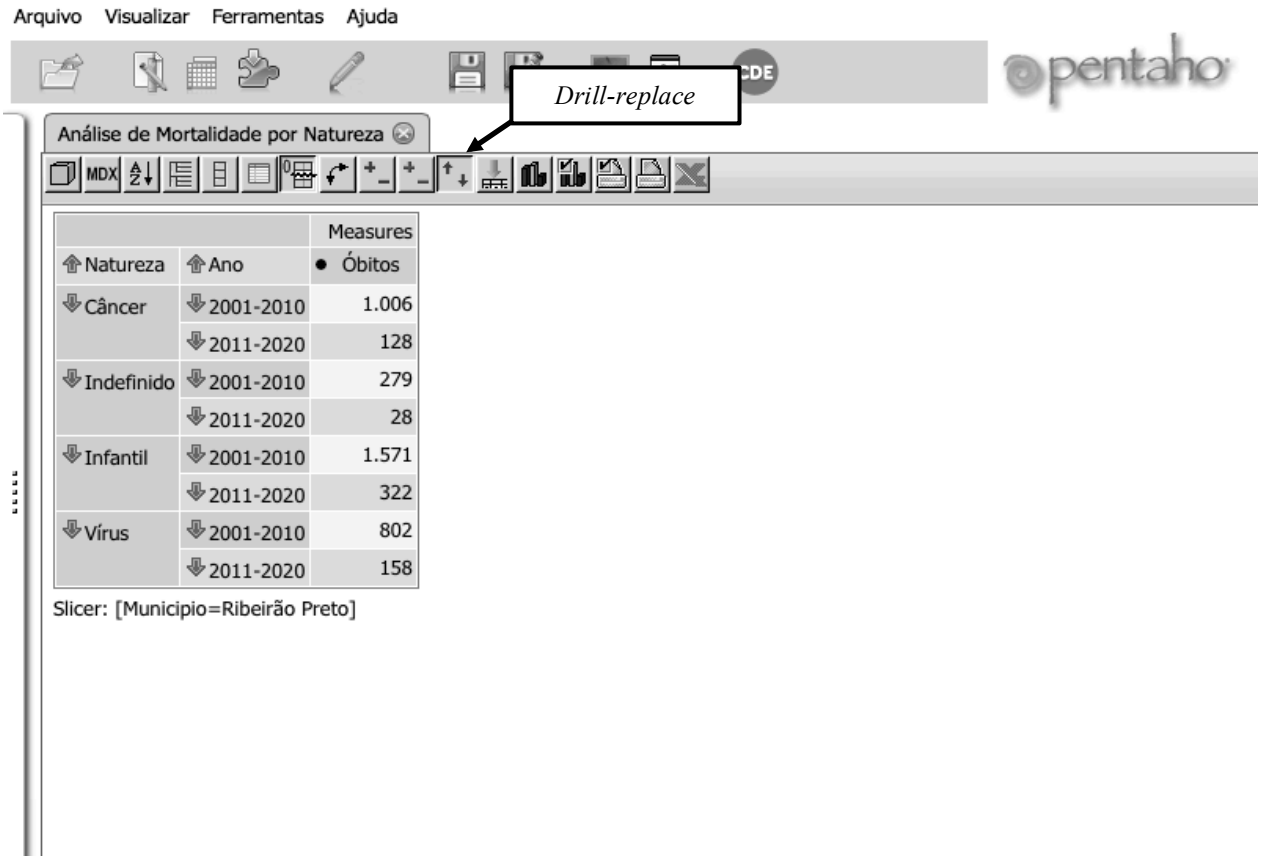


Figura 45 – Exemplo de consulta com a operação *drill-replace*.

Na Figura 45, podemos notar que os níveis expandidos foram os mesmos do exemplo anterior, entretanto, o resultado é diferente. Com a operação *drill-position* o primeiro passo desta consulta foi clicar no nível "Todas Natureza" da coluna "Natureza", depois de clicado, o nível "Todas Natureza" foi omitido na tabela e seus subníveis foram exibidos no lugar. O segundo passo desta consulta foi clicar no nível "Todos" da coluna "Ano", depois de clicado, o nível "Todos" foi omitido na tabela e seus subníveis foram exibidos no lugar.

Também podemos aplicar filtros de acordo com diferentes condições, os filtros tem função similar à operação *slice-and-dice* das operações OLAP. A Figura 46 apresenta um exemplo de aplicação de filtro na análise do cubo "Mortalidade_Natureza".

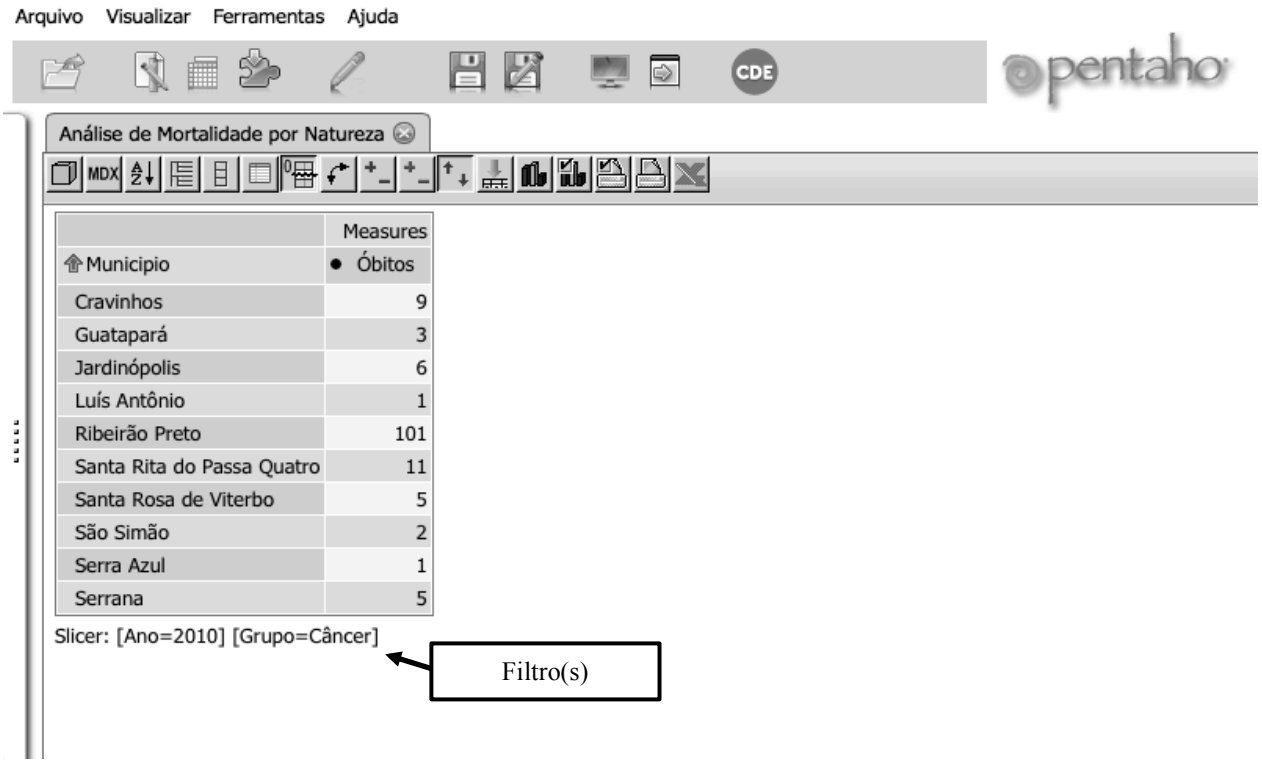


Figura 46 – Exemplo de aplicação de filtro na análise do cubo "Mortalidade_Natureza".

Na tabela resultante apresentada na Figura 46, podemos aplicar a operação de *swap axes* que é semelhante ao *pivot* das operações OLAP, a qual inverte os eixos (linha por coluna) da tabela. Além disso, também podemos configurar e mostrar um gráfico, de diferentes tipos (por exemplo barras, pizza, linhas, área, etc.), sobre os resultados destas operações. A Figura 47 apresenta um exemplo da operação *swap axes* e a geração de um gráfico de pizza por colunas sobre os resultados.

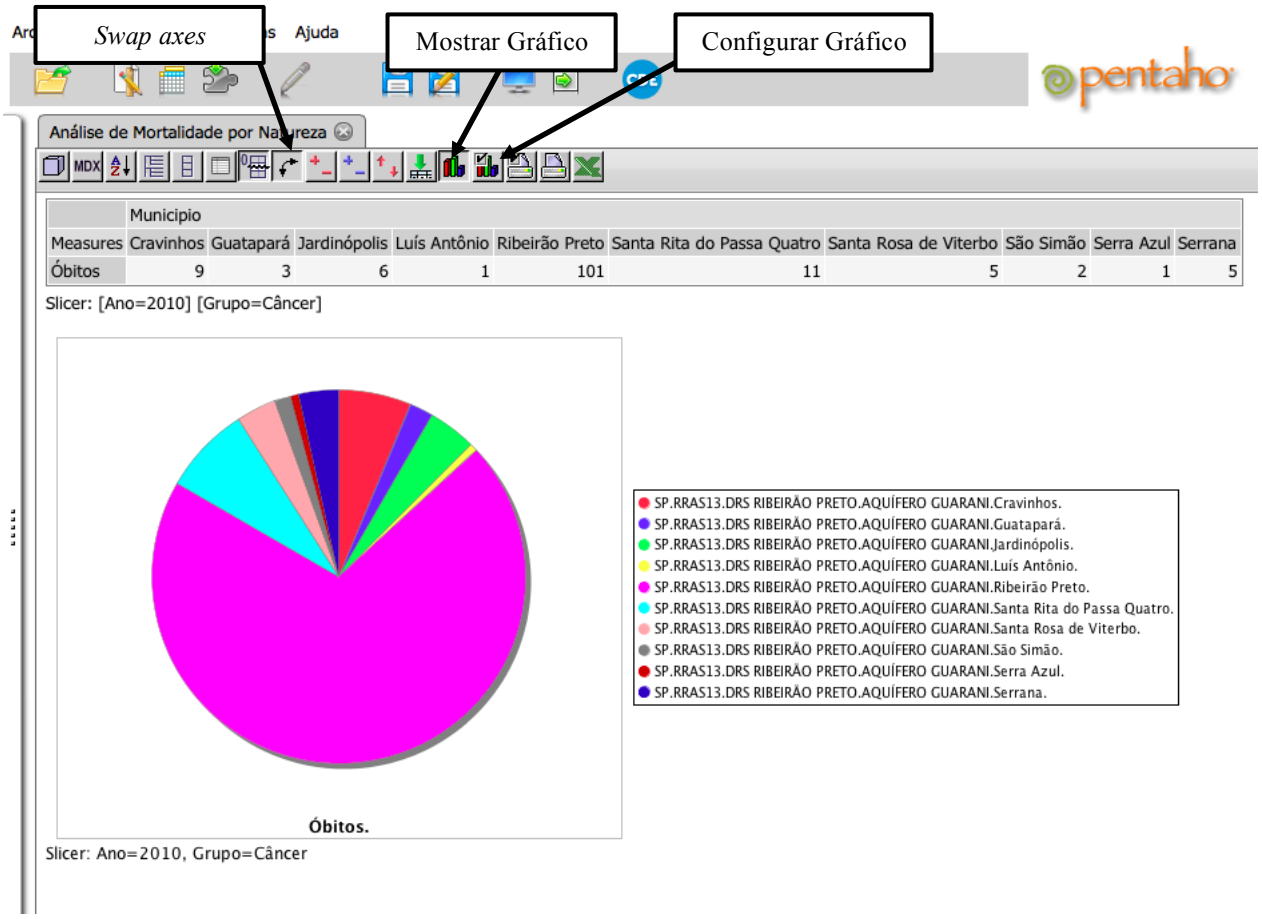


Figura 47 – Exemplo da operação *swap axes* e de criação de gráfico.

Além destas operações que podem ser realizadas sobre a tabela dinâmica do JPivot, também podemos exportar estas consultas para o formato XLS, da Microsoft Excel, e para o formato PDF. Os três últimos botões da Figura 48 são utilizados, respectivamente, para configurar a impressão em PDF, para realizar a impressão em PDF e para exportar em XLS.

Figura 48 – Menu de opções do JPivot.