

UNIVERSIDADE DE CAXIAS DO SUL
Centro de Computação e Tecnologia da Informação
Curso de Bacharelado em Ciência da Computação

Graziela Bettoni

ESTUDO DE CASO SOBRE UMA FERRAMENTA *OPEN SOURCE*
DE *BUSINESS INTELLIGENCE*
APLICADA À ÁREA DA SAÚDE

Caxias do Sul

2010

Graziela Bettoni

**ESTUDO DE CASO SOBRE UMA FERRAMENTA *OPEN SOURCE*
DE *BUSINESS INTELLIGENCE*
APLICADA À ÁREA DA SAÚDE**

Trabalho de Conclusão de Curso
para obtenção do Grau de
Bacharel em Ciência da
Computação da Universidade de
Caxias do Sul.

**Carine Geltrudes Webber
Orientadora**

Caxias do Sul

2010

AGRADECIMENTOS

A Deus e a Nossa Senhora de Caravaggio, por abençoar a mim e minha família, pois o que seria de nós sem a fé que possuímos neles.

Aos meus pais, Felix e Djanira e meu irmão, Diogo, que com muito carinho e apoio, não mediram esforços para que eu chegasse até esta etapa de minha vida.

Ao meu namorado, Almir, pela compreensão e dedicação ao longo deste dias.

A professora Orientadora, Carine, pela paciência na orientação, dedicação, sugestões, disponibilidade, credibilidade e incentivo que tornaram possível a elaboração e conclusão desta monografia.

A todos os professores da UCS, que foram importantes em minha vida acadêmica e no desenvolvimento desta monografia.

Aos amigos e colegas, pelo incentivo e apoio constante.

EPÍGRAFE

O verdadeiro sábio
é um eterno aprendiz.

Autor desconhecido

RESUMO

Business Intelligence (BI) é uma ferramenta que tem se tornado essencial em grandes organizações públicas ou privadas. Ela trabalha com dados armazenados em bancos de dados distintos, que são integrados em um *Data Warehouse* ou *Data Mart*. *Data Warehouse* é um banco de dados criado através da integração de diversos bancos heterogêneos. Ele tem a finalidade de trabalhar com um grande volume de dados e padronizá-los. O BI possibilita um diferencial competitivo, pois disponibiliza aos seus usuários uma visão mais correta sobre as informações das organizações e auxilia na tomada de decisões mais acertadas e com maior rapidez. Atualmente, muitas áreas têm necessidade do uso de BI porque trabalham com um grande volume de dados, muitas vezes incompletos e sem um padrão definido. Uma dessas áreas é a da saúde, principalmente no que diz respeito aos dados da saúde pública no Brasil. Em geral tais dados estão armazenados em bancos de dados distintos por serem oriundos de sistemas de informação distintos. Sem integração dos dados, não há como realizar a análise mais precisa sobre os indicadores de saúde da população brasileira. O alto custo da implantação de uma ferramenta BI pode ser um grande obstáculo na utilização desta tecnologia pela área da saúde. Com o surgimento de softwares de código aberto, a área da saúde pode dispor desta tecnologia. Uma dessas ferramentas existentes no mercado é o *Pentaho* (*Pentaho*, 2010), que possibilita ao usuário uma interatividade dinâmica na visualização dos dados. Neste trabalho, que está inserido no projeto OTICSSS (Webber, 2009), esta ferramenta é utilizada juntamente com dados provenientes do sistema de informação do DATASUS, em um estudo de caso. O estudo tem como objetivo verificar e analisar a viabilidade de implantação de ferramentas de BI na área da saúde. Como resultado deste trabalho, destaca-se a viabilidade no uso desta tecnologia e a capacidade de gerenciamento de grandes volumes de dados demonstrada pela ferramenta *Pentaho*.

Palavras-chaves: Área da Saúde. *Business Intelligence*. *Pentaho*. Banco de Dados. Integração de Banco de Dados. *Data Warehouse*. *Data Mart*.

ABSTRACT

Business Intelligence (BI) is an implement that has been essential in big organizations either public or private. It works with data stored in distinct data warehouse, which are part of *Data Warehouse* or *Data Mart*. *Data Warehouse* is a database created through the integration of many medley bench. Its aim is work with a large volume of data and also patterned them. BI enables a competitive increment, because enables to its users a better vision about the organizations information and helps in faster and precisely decision taking. Nowadays many areas need the implement of a BI because they work with a large number of data, usually incomplete and without a defined pattern. One of these areas is the health service, mostly the Brazil`s health service data. In general such data are stored in distinct database because they are originally from distinct information system. Without the data integration, it is not possible to make a precisely analysis about the Brazilian population`s health. The higher cost of an implement installation such as BI can be a big obstruction in the using of this technology by the health service. With the beginning of open source softwares the health service can use this technology. One of these implement is the *Pentaho* (*Pentaho*, 2010) which enables the user a dynamic interactivity on the data view. On this work that is inserted in the OTICSSS (Webber, 2009) project this implement is used along with data coming from the DATASUS information system, in a study of case. The study aims verifying and analyzing the availability of the BI implement installation in the health service area. As result of this work stands the availability in the use of this technology and the capacity management of large data volume showed by the *Pentaho* implement.

Keywords: Health Service. *Business Intelligence*. *Pentaho*. Database. Database Integration. *Data Warehouse*. *Data Mart*.

LISTA DE FIGURAS

Figura 1: Uma arquitetura de alto nível de BI (Turban, 2009).....	15
Figura 2: O processo de ETL (Turban, 2009).....	18
Figura 3: Tabela de Fatos (Kimball, 1998).....	19
Figura 4: Tabela de dimensão (Kimball, 1998)	20
Figura 5: Representação de um Cubo de Dados (adaptação, <i>Microsoft</i> , 2010).....	22
Figura 6: Exibição do resultado das análises do cubo em formato de tabela	25
Figura 7: Exibição do resultado das análises do cubo em formato de gráfico	25
Figura 8: Arquitetura da Plataforma do BI <i>Pentaho</i> (adaptação, <i>Pentaho</i> , 2010).....	28
Figura 9: Processo ETL dos dados da tabela clientes para a tabela <i>tb_dimensao_clientes</i>	33
Figura 10: Criação do Cubo <i>produto_qtde</i>	33
Figura 11: Resultado do cubo <i>produto_qtde</i> através do <i>Pentaho Analysis</i> (<i>Pentaho</i> , 2010) ...	34
Figura 12: Modelagem multidimensional do Cenário I.....	44
Figura 14: ETL para tabela de dimensão tempo	48
Figura 15: ETL para tabela de dimensão local	49
Figura 16: ETL para Tabela de Dimensão Causa Básica.....	50
Figura 17: ETL para Tabela Fato de Mortalidade Causa Básica	51
Figura 18: <i>Schema</i> do Cenário I	53
Figura 19: Resultado exibido na ferramenta <i>Pentaho Analysis View</i>	55
Figura 20: Modelagem Multidimensional do Cenário II.....	56
Figura 21: ETL para Tabela Fato de Mortalidade Menores de 5 anos por Infecção Respiratória Aguda	59
Figura 22: ETL para Tabela Fato de Mortalidade Menores de 5 anos	60
Figura 23: <i>Schema</i> do Cenário II.....	62
Figura 24: Resultado exibido na Ferramenta <i>Pentaho Analysis View</i>	63
Figura 25: Barra de ferramentas do <i>Schema Workbench</i> (<i>Workbench</i> , 2010)	69
Figura 26: Barra de ferramentas do <i>Analysis View</i> (<i>Pentaho</i> , 2010).....	70
Figura 27: Arquivo XML Cenário I.....	78
Figura 28: Barra Arquivo XML Cenário II.....	80

LISTA DE TABELAS

Tabela 1: Tabela dos indicadores do Grupo C, da RIPSA.....	40
Tabela 2: Cadastros de Estados	42
Tabela 3: Cadastro de Grupos de Causas Básicas por Doenças do Aparelho Circulatório	43
Tabela 4: Estrutura da Tabela btbdimensaotempo	45
Tabela 5: Estrutura da Tabela btbdimensaolocal	45
Tabela 6: Estrutura da Tabela tbdimensaocausabasica	46
Tabela 7: Estrutura da Tabela ctbfatomortalidadecausabasica	47
Tabela 8: Estrutura da Tabela tbfatomortalidademenores5anosinfeccaoespiratoriaAguda_IndicadorC7	57
Tabela 9: Estrutura da Tabela tbfatomortalidademenores5anos	58
Tabela 10: Tabela das Causas Básicas por Doenças do Aparelho Circulatório.....	81

LISTA DE ABREVIATURAS E SIGLAS

Sigla	Significado
BI	<i>Business Intelligence</i>
CID10	10ª Revisão da Classificação Internacional de Doenças
CRM	<i>Customer Relationship Management</i>
DATASUS	Departamento de Informática do Sistema Único de Saúde
DN	Declarações de Nascimentos
DO	Declarações de Óbito
DW	<i>Data Warehouse</i>
EIS	Sistemas de Informações Executivas
ERP	<i>Enterprise Resource Planning</i>
ETL	Extração, Transformação e Carga
IBGE	Instituto Brasileiro de Geografia e Estatística
IDB	Indicadores e Dados Básicos para Saúde
IPEA	Instituto de Pesquisa Econômica Aplicada
JRE	<i>Sun Java Runtime Environment</i>
MDX	<i>Multidimensional Expressions</i>
MPL	<i>Mozilla Public Licence</i>
OLAP	Processamento Analítico Online
OTICSSS	Observatório de Tecnologias de Informação e Comunicação em Sistemas e Serviços de Saúde
PDI	<i>Pentaho Data Integration</i>
PPL	<i>Pentaho Public Licence</i>
RIPSA	Rede Interagencial de Informações para a Saúde
SIM	Sistema de Informações de Mortalidade
SINASC	Sistema de Informações de Nascidos Vivos
SUS	Sistema Único de Saúde
TI	Tecnologia da Informação

SUMÁRIO

1	INTRODUÇÃO	12
2	<i>BUSINESS INTELLIGENCE</i>.....	14
2.1	<i>Data Warehouse</i>	15
2.2	<i>Data Mart</i>	16
2.3	Extração, Transformação e Carga	17
2.4	Modelagem Dimensional	18
2.4.1	Tabela de Fatos	19
2.4.2	Tabela Dimensão	20
2.4.3	Tabelas Agregadas	20
2.4.4	Cubo de Dados	21
2.5	Mineração de Dados	22
2.6	Processamento Analítico <i>Online</i>	22
2.7	Visualização dos Dados	24
2.8	Considerações Finais	26
3	<i>PENTAHO</i>.....	27
3.1	Ferramenta <i>Open Source</i> de BI <i>Pentaho</i>	27
3.2	Necessidades da Ferramenta	28
3.3	<i>Pentaho</i>	29
3.3.1	<i>Pentaho Data Integration</i>	29
3.3.2	<i>Mondrian/JPivot</i>	30
3.3.3	<i>Schema Workbench</i>	31
3.3.4	<i>Analysis View</i>	31
3.3.5	<i>Report Designer</i>	32
3.4	Exemplo de Uso	32
3.5	Considerações Finais	35
4	BI NA ÁREA DA SAÚDE.....	36
4.1	Importância do BI na área da Saúde	36
4.2	Origem dos Dados	37
4.3	Indicadores na Área da Saúde.....	38
4.4	Visão Geral da Implementação	39

4.5	Cenário I: Grupo de Causas por Doenças do Aparelho Circulatório	43
4.5.1	Modelagem	44
4.5.2	Processo de <i>ETL</i>	47
4.5.3	Implementação.....	52
4.5.4	Visualização.....	54
4.6	Cenário II: Indicador C.7 - Mortalidade Proporcional por Infecção Respiratória Aguda em Menores de 5 Anos	55
4.6.1	Modelagem	55
4.6.2	Processo de ETL.....	58
4.6.3	Implementação.....	61
4.6.4	Visualização.....	63
4.7	Considerações Finais	63
5	CONSIDERAÇÕES FINAIS	65
5.1	Conclusão.....	65
5.2	Trabalhos Futuros	66
6	REFERÊNCIAS	67

1 INTRODUÇÃO

As organizações em geral precisam ter atitudes rápidas, eficientes e inovadoras para sobreviverem e se destacarem no mercado de trabalho. Essas atitudes exigem conhecimento e estratégias dos gestores das empresas, tornando as decisões a serem tomadas mais complexas, devido à manipulação do crescente volume de dados. As decisões devem ser baseadas em uma quantidade elevada de dados adequados e relevantes, além de conhecimentos e informações dos gestores da empresa. Para que esse processo seja feito com eficiência, agilidade e clareza, é necessário a utilização de ferramentas de apoio para a tomada de decisões. Estas ferramentas fazem parte da categoria de softwares de inteligência de negócio, ou *Business Intelligence* (BI), as quais têm a finalidade de coletar, analisar e distribuir os dados de uma forma melhor e mais apropriada para auxiliar nas tomadas de decisões (Turban, 2009).

Ferramentas de BI se baseiam na utilização de fontes variadas de informação para se definir estratégias de competitividade nos negócios da empresa. Elas compreendem um conjunto de ferramentas, usadas para trabalhar com um grande número de dados em busca de informações mais precisas e claras para empresa. Uma análise profunda dos dados, informações e conhecimentos da organização permite um crescimento na mesma e eficiência na competitividade (Barbieri, 2001).

No mercado de trabalho existem diversas ferramentas de BI e/ou técnicas que buscam atender às necessidades das organizações que almejam obter vantagens competitivas. Todas estas ferramentas possuem um fundamento principal, que consiste na extração e análise de informações significativas e úteis para a tomada de decisões. As informações são geradas através dos negócios e interesses da organização, resultando em um grande volume de dados. Algumas das ferramentas de BI utilizadas para esta finalidade são: *Database marketing*, *Data Mart*, *Data Warehouse*, *OLAP (On Line Analytical Processing)*, *Mineração de Dados*., *Webhouse* (Serra, 2002).

Existem alguns softwares *open source* de BI no mercado como: *Pentaho* (Pentaho, 2010), *Spago BI* (SpagoWorld, 2010) e *Jaspersoft BI Suite* (Jaspersoft, 2010), sendo que a tecnologia de BI pode ser utilizada visando transformar um grande volume de dados em informações qualitativas. O *Pentaho* fornece relatórios diversos, análises de dados e

minerações de dados, sendo a ferramenta pioneira no mercado *open source* de BI, tendo uma ótima aceitação nas implementações em empresas e organizações (Pentaho, 2010).

Neste trabalho propõe-se o uso de ferramentas de BI para o caso de tratamento de dados da área da saúde. As organizações públicas e que atuam na área de saúde, normalmente têm os cadastros das informações em banco de dados relacional e não há integração entre sistemas, pois cada um preenche as informações em bancos de dados distintos. Além disso, no Brasil não são aplicados padrões internacionais para a modelagem dos processos e dados.

Este trabalho desenvolve uma análise da ferramenta de software de código aberto *Pentaho* aplicada à área da saúde. Ele tem a finalidade de testar os recursos que podem ser utilizados para avaliar as necessidades e critérios específicos identificados pelos profissionais da área da saúde, para uma melhor visualização e interpretação das informações.

Para o estudo do uso de ferramentas de BI para auxiliar o tratamento e acesso aos dados da área da saúde, o trabalho está dividido em 5 capítulos: o capítulo 2 descreve uma fundamentação teórica sobre a tecnologia de BI, seus principais conceitos e padrões para uma melhor compreensão do trabalho.

O capítulo 3 apresenta a ferramenta de BI *open source Pentaho*, seus recursos, funcionalidades e um exemplo de uso.

O capítulo 4 descreve a origem e a base de dados da área da saúde utilizada como fonte de informação, a fim de identificar os critérios específicos da área que possam ser aplicados na integração e análise dos dados na ferramenta *open source* de BI *Pentaho*. Além disso, é apresentado o estudo de caso ilustrando os processos e etapas de dois cenários. E no final são explanados pontos importantes encontrados na execução do estudo de caso.

O capítulo 5 apresenta as conclusões referente a realização do trabalho em um todo, bem como sugestões para trabalhos futuros.

2 ***BUSINESS INTELLIGENCE***

Nos anos 70, os relatórios eram estáticos, bidimensionais e não tinham como ser analisados automaticamente. No início da década seguinte surgiu o que se chamou de Sistemas de Informações Executivas (EIS). Com eles apareceram mais recursos como relatórios dinâmicos, multidimensionais, onde os gerentes ou executivos conseguiam fazer uma análise mais aprimorada e precisa das informações. Através desta análise, conseguiam extrair tendências e fatores críticos para obter um melhor aproveitamento a fim de atingir o sucesso. Após os meados de 1990 começaram aparecer estas tecnologias com o nome de *Business Intelligence*. Em 2005, começaram aparecer recursos maiores de análises e incorporou-se a utilização de inteligência artificial nos sistemas que utilizam tecnologia de BI. Atualmente, se reconhece que um bom sistema que utilize informações gerenciais é baseado em recursos de BI. As ferramentas mais sofisticadas incluem a grande maioria dos recursos de BI e outras se especializam em apenas alguns destes recursos (Turban, 2009).

O termo BI foi estabelecido pelo *Gartner Group* (Gartner, 2010). BI é composto por várias ferramentas que colecionam e armazenam os dados, fazem integração, mineração e análises deles. Nelas são utilizadas diversas tecnologias para visualização, com a finalidade de gerar relatórios e mostrar os resultados com informações significativas, para que se consiga tomar uma decisão mais precisa (Dayal, 2009). BI, de uma maneira geral, é compreendido como a utilização de variadas fontes de informação para se definir estratégias de competitividade nos negócios da empresa. Ele é um conjunto de ferramentas usadas para trabalhar com um grande número de dados operacionais em busca de informações mais precisas e claras para o negócio. Estas ferramentas utilizam-se dos dados, informações e conhecimentos permitindo uma maior competitividade eficiente entre as organizações (Barbieri, 2001).

A arquitetura de BI foi projetada para gerar informações estratégicas para tomada de decisões, onde alguns especialistas analisam um grande volume de dados, a fim de construir modelos e preparar relatórios para um período de tempo (Dayal, 2009). Um BI é composto por *Data Warehouse*, mineração de dados, análise dos dados e visualização das informações.

A figura 1 ilustra uma arquitetura de alto nível de BI, composta por quatro grupos: Ambiente de *Data Warehouse*: o *Data Warehouse* carregado com sua fonte de dados sobre a

análise de negócios da organização; Ambiente de análise de negócio: um conjunto de ferramentas para análise e manipulação dos dados contidos no *Data Warehouse*; Desempenho e estratégia: verificação e controle do desempenho dos resultados; e interface de usuário: modo de acesso do usuário final (Turban, 2009).

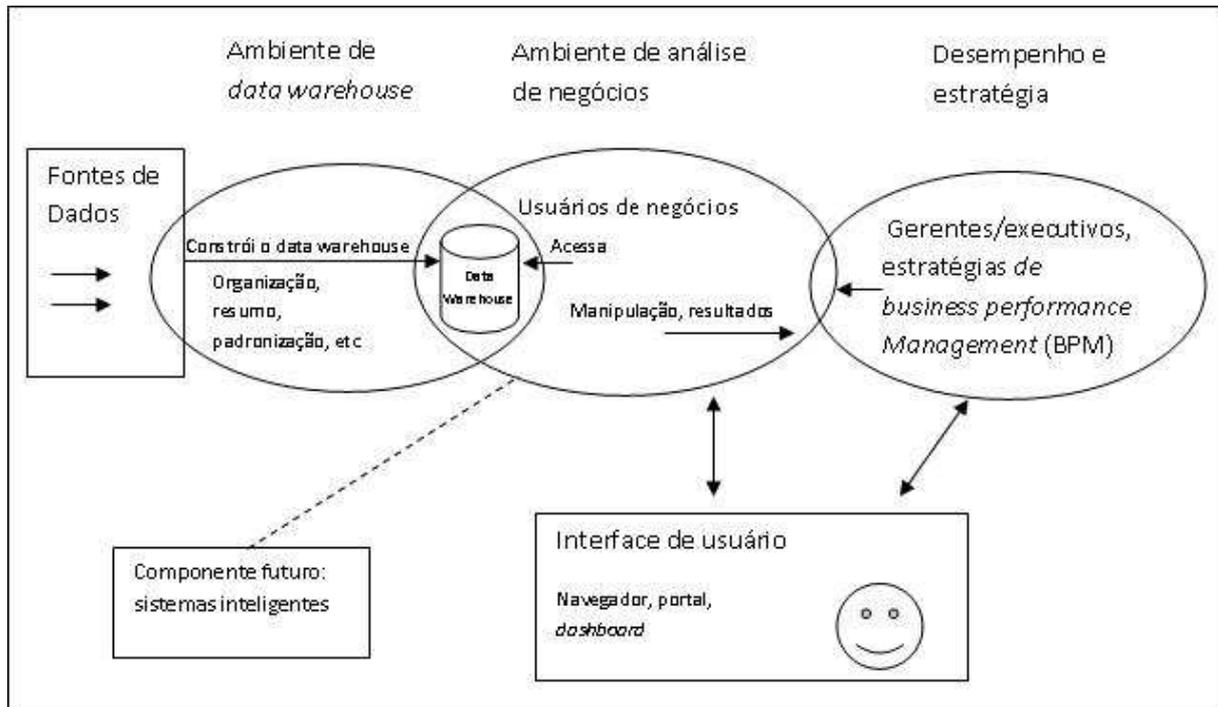


Figura 1: Uma arquitetura de alto nível de BI (Turban, 2009)

2.1 *Data Warehouse*

Segundo Inmon (1997), a definição de *Data Warehouse* (DW) é “um conjunto de dados baseados em assuntos, integrado, não-volátil, variável em relação ao tempo, de apoio às decisões gerenciais”. Assim sendo, nos DWs são gravadas informações conforme um grupo de assunto das atividades de negócio, ou seja, os principais assuntos que são de interesse da organização, proporcionando uma visão maior. A integração de dados é responsável por analisar os dados que já existem em outros sistemas da empresa e torná-los padrões e consistentes para que possam ser utilizados enfrentando a diversidade de formatos encontrados. A não-volatilidade se dá porque os dados são armazenados e consultados, onde estes dados geralmente não sofrem alterações e a variável de tempo está associada a um ponto

no tempo para compor os dados históricos.

No *DW* mais duas características são importantes salientar, a granularidade e o particionamento dos dados. A granularidade é o nível de detalhe utilizado para a criação de um *DW*, isto atinge diretamente o armazenamento da grande quantidade de armazenamento e também o tipo das consultas que possam ser respondidas. Quanto maior o nível de granularidade, menor a quantidade de índices e dados terá, e conseqüentemente também uma menor possibilidades de consultas detalhadas. Porém, com um nível de granularidade baixa podem ser feitas consultas para questões mais específicas, em contrapartida é necessário recursos computacionais maiores. A maioria das organizações gostaria de obter eficiência nos dois pontos positivos, ou seja, um grande volume de armazenamento de dados e com consultas detalhadas, entretanto, para que isso ocorra é necessário um equilíbrio, um balanço no nível que será utilizado, tornando-se quase padrão a utilização de duplos níveis de granularidade (Serra, 2002).

Um nível duplo de granularidade se refere a uma camada para os dados que são armazenados na área operacional, estruturados resumidamente para utilização dos gestores e analistas da organização, e outra camada para os dados de contexto históricos com os detalhes. Neste segundo nível os dados podem ser armazenados em outra unidade física do computador (Serra, 2002). O particionamento dos dados em mais de uma unidade física torna os processos mais flexíveis. Isso se deve ao grande volume de dados que podem ser armazenados em um *DW* e conseqüentemente a demora que isso pode causar nos processos.

2.2 Data Mart

Data Mart é um subconjunto de um *DW* contendo os dados importantes para uma área e/ou departamento da empresa. O *Data Mart* pode servir para testes quando há a intenção de utilizar um *DW* e assim verificar sua eficiência e resultados em um escopo menor. Normalmente, pode-se começar desenvolvendo um *Data Mart* e após ir ampliando aos poucos o processo. O principal elemento para o sucesso de um *Data Mart* ou até mesmo de um *DW*, estão nos tipos de informações valiosas que são identificadas pelas organizações. O processo de criação de um *DW* deve ser gradual ao longo do tempo. As diferenças que existem entre um *Data Mart* e *DW* são o tamanho e o foco do problema a ser resolvido. Um *Data Mart* trata de um problema setorial ou local, enquanto que um *DW* abrange todos os

níveis da organização e conseqüentemente este tem um custo e uma demora maior na elaboração e desenvolvimento (Serra, 2002).

2.3 Extração, Transformação e Carga

Segundo TURBAN (2009), o processo de Extração, Transformação e Carga (ETL) consiste em: *extração*, onde é feita a leitura dos dados de um ou vários bancos de dados operacionais; *transformação*, onde são formatados e convertidos os dados de um determinado banco origem conforme padronização utilizada no banco de dados de destino, seja ele um *DW* ou até mesmo outro banco; e *carga*, onde são colocados os dados dentro do *DW*. Nesta fase também são padronizadas as chaves primárias e estrangeiras, a divisão nas unidades físicas e a criação dos índices. A transformação acontece através de tabelas de busca ou regras, ou combinação dos dados com outros dados.

O objetivo do ETL é a integração e limpeza dos dados. Os dados podem ser originados de qualquer fonte, como por exemplo: uma aplicação ERP (*Enterprise Resource Planning*), uma ferramenta de CRM (*Customer Relationship Management*), um arquivo texto, entre outros (Turban, 2009).

A figura 2 apresenta o processo de ETL. Ele consiste em: extrair dados de uma aplicação pronta, de um sistema legado ou de outras aplicações internas da organização; depois fazer a transformação e limpeza dos dados e por fim carregar em um *Data Mart* ou *DW*.

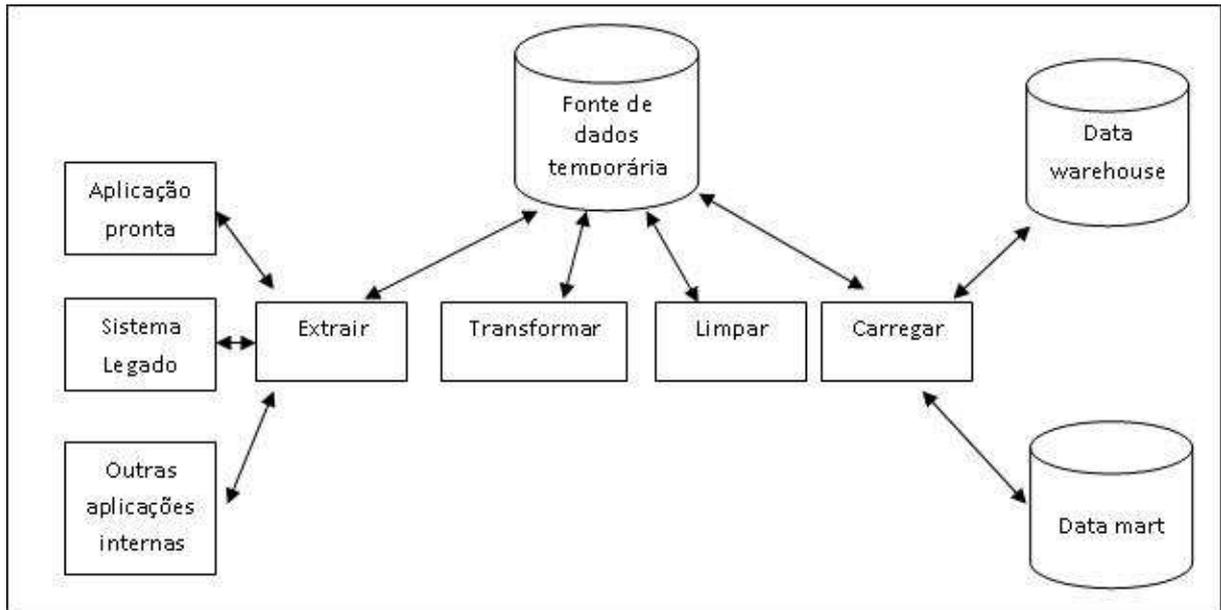


Figura 2: O processo de ETL (Turban, 2009)

As tecnologias de ETL são executadas para o uso e processo de *DWs*. Este é um processo de suma importância para qualquer projeto que tenha um foco principal em dados. Os gerentes de TI enfrentam vários desafios, pois frequentemente se ocupa 70% do tempo trabalhando com ETL em um projeto focado em dados (Turban, 2009).

O ETL apresenta um desempenho adequado e uma funcionalidade correta no mapeamento dos dados distribuídos em diversos setores da organização. Apesar da quantidade de informações, para que estes dados sejam integrados com uma alta qualidade será necessária uma visão aprimorada e mais eficiente. Este processo envolve diversos riscos devido à quantidade enorme de dados que são trabalhados e a diversidade de bancos que podem ser encontrados (Dayal, 2009).

2.4 Modelagem Dimensional

A modelagem dimensional é uma técnica que organiza e estrutura os dados para que sejam consultados de uma maneira que gerem as informações de maneira intuitiva, podendo ser vistas de perspectivas diferentes, facilitando a análise pelos gestores das organizações (Kimball, 1998; Machado, 2000).

Neste tipo de modelagem, os dados referentes ao negócio da organização são

estruturados em mais de uma dimensão, e a intersecção destas dimensões resulta normalmente em valores numéricos que representam as medidas ou fatos. O resultado deste tipo de modelagem é um modelo conceitual com dimensões, o qual é formado por tabelas Fatos e tabelas Dimensão. Ele gera uma visão mais simples de interpretação.

2.4.1 Tabela de Fatos

Fato é um assunto referente aos negócios da organização e possui valores numéricos para representá-lo. Além disso, seus valores podem ser alterados e analisados com o passar do tempo. Diversos trabalhos aplicados exploram este conceito (Machado,2000; Serra, 2002; Barbieri, 2001). Tabela de Fatos é uma tabela que compõe diversos registros que correspondem aos resultados dos fatos que serão analisados. Ela possui as chaves primárias das tabelas de Dimensão, responsáveis pela conexão com as mesmas, e contém as informações para a geração dos relatórios a fim de visualizar as consultas (Turban, 2009). Este tipo de tabela pode conter campos que podem ser sumarizados e que são necessários para obtenção de informações significativas para análise.

A figura 3 ilustra a Tabela de Fato Vendas, formada pelas chaves estrangeiras: chave_tempo, chave_produto e chave_loja que compõe a chave primária desta tabela. Além disso, é composta por mais três medidas: vendas_em_dolar, unidades_vendidas e custos_em_dolar.

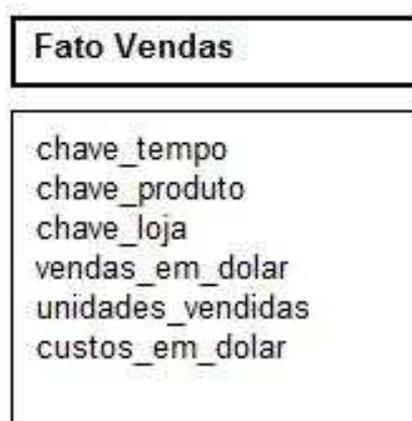


Figura 3: Tabela de Fatos (Kimball, 1998)

2.4.2 Tabela Dimensão

Dimensões são as entidades que participam de um fato. Diversos trabalhos aplicados exploram este conceito (Machado,2000; Serra, 2002; Barbieri, 2001). Tabela Dimensão é constituída pelos registros de entrada dos dados. Ela tem uma relação de um para muitos com a tabela Fato e possui várias colunas de informações. As dimensões são as tabelas a partir das quais a tabela Fato busca as descrições e informações das chaves contidas nela. Elas possuem uma chave primária que é ligada à tabela Fato (Barbieri, 2001).

A figura 4 ilustra as tabelas de dimensão: Tempo, Produto e Loja, cada uma com sua chave primária, sendo que na tabela de fatos elas são chaves estrangeiras.

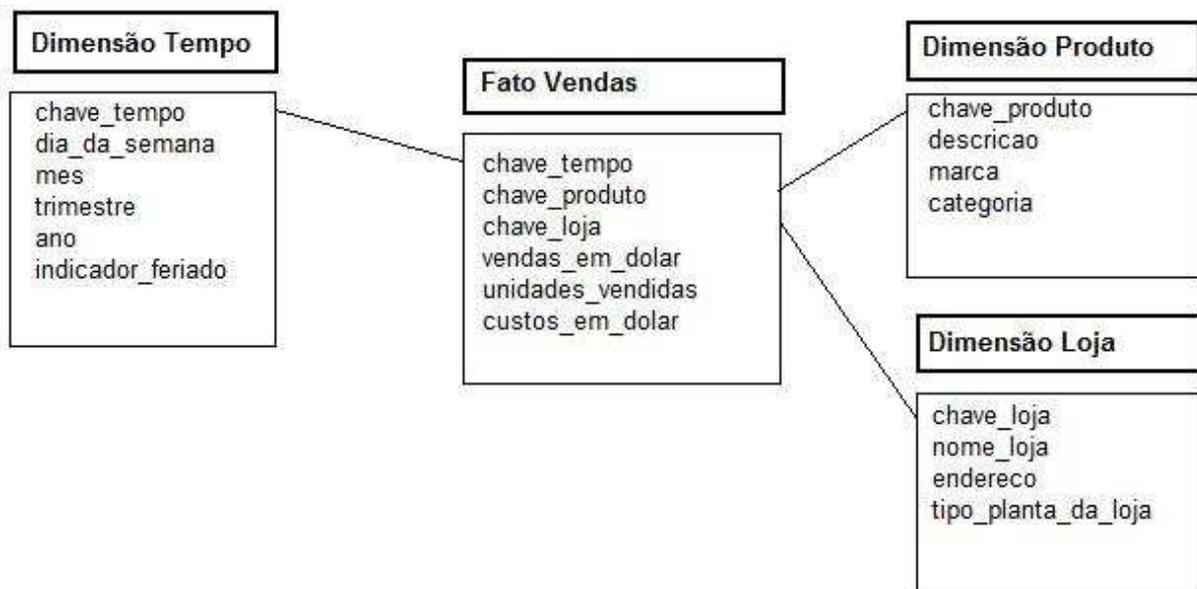


Figura 4: Tabela de dimensão (Kimball, 1998)

2.4.3 Tabelas Agregadas

As tabelas agregadas são criadas, quando necessário, com a finalidade de melhorar o desempenho das consultas resultantes para análise dos dados. Estas consultas resultantes podem se deparar com um volume enorme de dados ocasionando uma demora elevada na criação da carga inicial de dados e na sua leitura. Estas tabelas são criadas através da

granularidade das tabelas fatos, gerando tabelas menores e com a soma de seus resultados (Machado, 2000).

Este tipo de tabela pode gerar vantagens e desvantagens. A vantagem é uma maior rapidez de acesso as consultas e ao processo de tomada de decisão, pois estará acessando tabelas com os dados resumidamente. Em contrapartida, existe um aumento na utilização de espaço de armazenamento, pois pertencerão a uma coleção de tabelas Fato e Dimensão e afetarão a premissa da não-redundância nos bancos de dados (Barbieri, 2001).

2.4.4 Cubo de Dados

O cubo é uma figura representativa de um conjunto de dados de fatos e de dimensões, que são extraídos de uma parte do *DW* ou de um *Data Mart*, a fim de atender as necessidades específicas ou alguma medida de interesse (Barbieri, 2001). O cubo de dados é o desenho que geralmente ilustra um modelo dimensional. Um modelo com três dimensões é possível ser representado através de um cubo. Na maioria dos casos é utilizado mais que três dimensões para formar um modelo dimensional, ao qual se dá o nome de Hipercubo. A visualização destes tipos de modelos é mais difícil. Qualquer modelo multidimensional é referenciado a um cubo (Machado, 2000).

Cada dimensão do cubo se refere a um atributo no banco de dados e cada célula encontrada neste cubo é a representação da necessidade de interesse. Os cubos são modelos eficientes de obter resultados mais precisos para os responsáveis pela tomada de decisão, através de análises e consultas. A interpretação do cubo de dados é uma maneira rápida e eficiente de gerar consultas e relatórios para os usuários finais (Turban, 2009).

A figura 5 esboça um Cubo de dados, onde as arestas representam as Dimensões, (nesta figura são: Tempo, Região e Produto) e cada célula representa um indicador do resultado do cruzamento de uma ou mais dimensões.

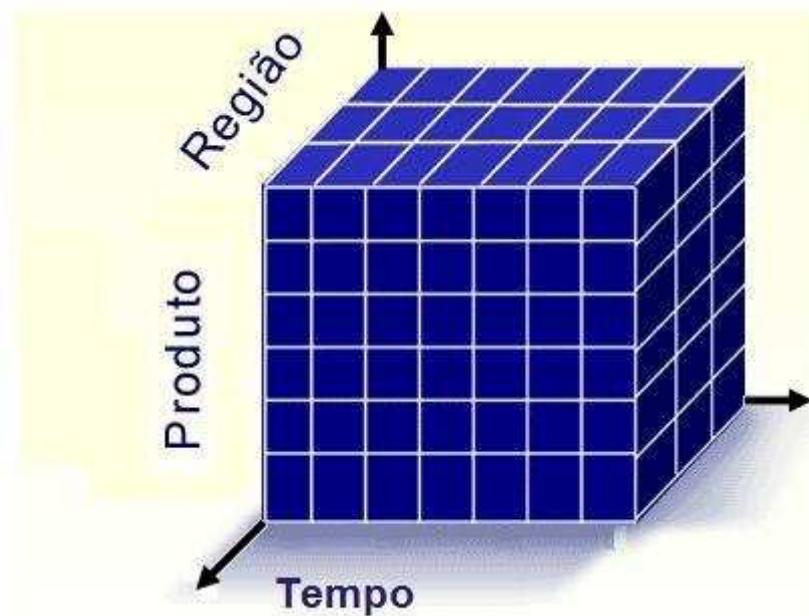


Figura 5: Representação de um Cubo de Dados (adaptação, Microsoft, 2010)

2.5 Mineração de Dados

A mineração de dados consiste em descobrir regras, tendências e modelos através de análises de um grande volume de dados, classificados para que os gestores possam visualizar informações, interpretar e tomar decisões com uma maior precisão (Fayyad, 1996; Waltz, 1999; Mitchell, 1997). A mineração de dados é feita através de algoritmos matemáticos, princípios da inteligência artificial, cálculos probabilísticos e combinações de diversos métodos estatísticos para extrair informações significativas para que se possa responder qualquer questão associada ao negócio (Delmater, 2001; Serra, 2002).

A mineração de dados tem um objetivo importante para os gestores de uma organização com a finalidade de conseguir apontar chances ou oportunidades de negócios, criando um diferencial competitivo sustentável (Delmater, 2001).

2.6 Processamento Analítico *Online*

Processamento Analítico *Online* (*Online Transaction Processing* ou *OLAP*) compreende um grupo de funções que auxiliam a análise multidimensional dos dados, através

de manipulação dos dados que são adicionados nas diversas categorias ou dimensões (Inmon, 1997). As ferramentas OLAP cruzam os dados de uma forma simples e eficiente, além de exibir resultados mais intuitivos, gerando informações mais precisas e concretas que facilitem a tomada de decisão pelos responsáveis da organização (Thomsen, 2002).

As consultas OLAP, normalmente, são acessadas por usuários finais, as quais incluem atividades como solicitação de relatórios, visualização de gráficos, construção e geração de respostas das consultas (Turban, 2009).

Diversas operações podem ser executadas para realizar a análise de um cubo no modelo multidimensional, gerando perspectivas, visões, informações e análises aperfeiçoadas. Exemplos de operações básicas de OLAP são:

Os tipos de operações *Drill-Down* e *Drill-Up* trabalham com o nível de granularidade e níveis hierárquicos entre as dimensões (Barbieri, 2001; Machado, 2000). *Drill-Down* é a operação de estar em um nível superior e se direcionar para um nível mais abaixo (Barbieri, 2001; Machado, 2000). Utilizado quando o usuário deseja ter informações mais detalhadas, diminuindo o grau de granularidade. *Drill-Up* é o processo de estar em um nível menor e se direcionar para um nível mais alto. Esta operação é a inversa da *Drill-Down* (Barbieri, 2001; Machado, 2000), sendo aplicada quando o usuário deseja ter um grau menor de detalhamento das informações, aumentando o grau de granularidade.

Drill-Across é a operação que você pula de um nível intermediário para outro nível intermediário. O nível deve estar com as dimensões em conformidades ou compartilhadas (Barbieri, 2001).

Drill-Through é uma operação que permite buscar informações de um nível de hierarquia menor do que consta na Tabela Fato, desde que a granularidade seja possível (Barbieri, 2001).

Slice and Dice serve para alterar a posição de uma informação, alterar colunas por linhas com a finalidade de facilitar a compreensão e fazer giros com o cubo. *Slice* é a operação de corte do cubo, mantendo a mesma perspectiva de demonstração dos dados. Ou seja, é o fatiamento do cubo. *Dice* é a alteração da perspectiva da visão (Machado, 2000).

Sorting é uma operação de ordenação crescente ou decrescente de uma determinada medida escolhida para visualização. Esta ordenação, apenas afeta o relatório de visualização não afetando em nada na consulta.

Consultas *ad-hoc* são consultas realizadas pelos usuários após a geração de relatórios,

para que sejam respondidas as questões que estão sem respostas pelo tomador de decisão. Estas consultas são realizadas através de uma estrutura ou método que não foi utilizado anteriormente, mas que o sistema entenda o que o usuário necessita (Turban, 2009).

2.7 Visualização dos Dados

Visualização dos dados é a apresentação visual dos resultados das consultas efetuadas no *DW* ou *Data Mart*. As tecnologias visuais têm como objetivo mostrar o resultado para os usuários de uma maneira de mais fácil compreensão e de chamar mais a sua atenção. Algumas vezes, esta tecnologia pode se referir também a interpretação das informações e dados em algum determinado momento no decorrer de todo o processo (Turban, 2009). As ferramentas visuais incluem geração de relatórios, planilhas, gráficos, imagens digitais, representações dimensionais, sistemas geográficos, vídeos e animações. Estas ferramentas podem também auxiliar na identificação de relações entre as informações, podendo extrair tendências e previsões (Turban, 2009).

A visualização dos dados também habilita utilizar ferramentas que tenham recursos para disponibilizar estas informações na Web. O responsável por analisar as informações pode utilizar uma interface de navegador para fazer as análises de dados em tempo real, ao contrário de esperar vários relatórios para fazer comparações e relações entre eles. Com a utilização dos recursos das tecnologias visuais de análise, os profissionais da área conseguem identificar riscos e problemas que podem ter ocorrido no decorrer do tempo sem serem notados, devido à utilização dos métodos padrão (Turban, 2009).

A seguir são apresentados alguns modelos de resultados possíveis para visualização das informações. A figura 6 ilustra uma visualização em formato de tabela, podendo aproveitar os recursos da análise OLAP para distribuir os dados com o objetivo de atingir o resultado desejado. Na figura 7 pode-se visualizar em formato de gráfico as informações resultantes, demonstrando para o usuário final que existem vários recursos visuais disponíveis e que ele pode aproveitar o recurso da maneira que achar melhor.

		Year - Quarter Month								
		1996		1997		1998				
						Quarter 1		Quarter 2		
Category Name	Product Name	Ship Country	Ship Region	Ship City	Unit Price	Quantity	Unit Price	Quantity	Unit Price	Quantity
Beverages					2033,8	1842	5058,35	3996	3622	2381
Condiments					705,6	962	2211,85	2895	1037,4	936
Confections					1333,4	1357	3547,63	4137	2059,63	1795
Dairy Products					1857,6	2086	5070	4374	1632,1	1649
Grains/Cereals	Filo Mix				16,8	48	102,2	313	49	105
	Gnocchi di nonna Alice				212,8	96	1178	971	304	171
	Gustaf's Knäckebröd				16,8	6	184,8	209	63	108
	Ravioli Angelo				93,6	133	148,2	124	136,5	129
	Singaporean Hokkien Fried Mee				54,6	37	215,6	451	56	97
	Tunnbröd				21,6	105	82,8	287	45	176
	Wimmers gute Semmelknödel				106,4	124	425,6	281	266	158
	Total				522,6	549	2337,2	2636	919,5	944
Meat/Poultry					1294,3	950	3795,48	2189	1473,09	527
Produce	Longlife Tofu				40	141	54	116	10	20
	Manjimup Dried Apples	Argentina					53	7		
		Austria					53	120		
		Brazil	RJ		84,8	37				
			SP	Campinas			53	28		
				Total	84,8	37	53	28		
		Canada					53	30		5
		Denmark								
		France			42,4	20	106	35	106	7
		Germany			47,4	40	57	48	57	9
		Ireland								

Figura 6: Exibição do resultado das análises do cubo em formato de tabela

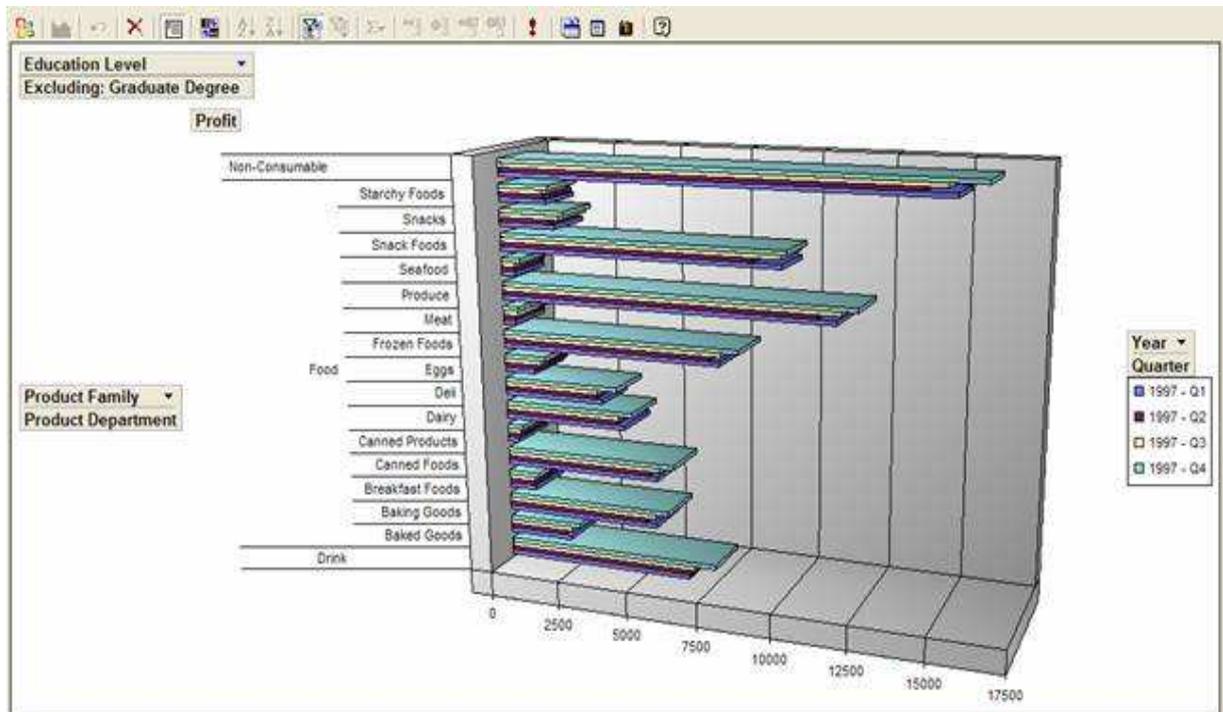


Figura 7: Exibição do resultado das análises do cubo em formato de gráfico

2.8 Considerações Finais

Este capítulo apresentou e explanou os conceitos fundamentais sobre a tecnologia de BI para uma melhor compreensão sobre o assunto. Esta tecnologia está crescendo e evoluindo cada vez mais, devido à manipulação de um grande volume de dados pelas organizações. No mercado de trabalho existem diversas ferramentas de BI, dentre elas e para este trabalho foi escolhida a ferramenta de BI *Pentaho*. O motivo da escolha é pela opção de utilizar um conjunto de ferramentas *open source* e por ele permitir a visualização do comportamento em relação a base de dados na área da saúde. No próximo capítulo são explicados os recursos e funcionalidades da ferramenta de BI escolhida, apresentando um exemplo de seu uso.

3 **PENTAHO**

Este capítulo apresenta a Ferramenta *open source* de BI *Pentaho* (Pentaho, 2010), seus recursos e funcionalidades, ilustrando por fim com um exemplo de uso.

3.1 **Ferramenta *Open Source* de BI *Pentaho***

Pentaho é um conjunto de ferramentas *open source* para criação de soluções de BI (Pentaho, 2010). Ele é composto por: *Pentaho BI Plataform*, *Pentaho Data Integration – PDI (Kettle)*, *Analysis View (Mondrian/JPivot)*, *Pentaho Reporting (Reporting)* e *Weka (Data Mining)*. Entre essas ferramentas, algumas já vêm integradas quando feito o download do *Pentaho*, para outras é necessário fazer *downloads* e a configuração do ambiente separadamente. Isto ocorre conforme a versão que é utilizada (Pentaho, 2010).

A plataforma de código aberto *Pentaho* é distribuída através da Licença Pública da *Pentaho (PPL – Pentaho Public Licence)*, *Mozilla Public Licence 1.1 (MPL 1.1)*, *GNU GPL* e *GNU Library*. O *Pentaho* é acessível através de *WebServices* e pode ser trabalhado tanto no ambiente *Windows* como *Linux* (Pentaho, 2010).

Os dados são localizados e coletados dos bancos relacionais, após são feitas as transformações e normalizações necessárias para o armazenamento nas tabelas: fatos, dimensões e agregadas. O resultado obtido é a modelagem multidimensional dos dados. O *Pentaho* não é capaz de identificar os cubos e dimensões somente com o modelo dimensional, é necessária a criação de um arquivo *XML* descrevendo essas informações. Além disso, a partir desse *XML* é preciso gerar um *XACTION*, que é o arquivo identificado pelo *Pentaho*. Normalmente, para esta etapa é utilizada a ferramenta *Schema Workbench*, pois tem uma integração com o *Pentaho* e faz a publicação do *XML* automaticamente. A análise do cubo criado é feito através de um Servidor *OLAP* chamado *Mondrian*, o qual utiliza como cliente o *JPivot*. O resultado final obtido é a geração de informações significativas e claras, as quais podem ser visualizadas em formatos de relatórios, gráficos ou tabelas (Pentaho, 2010).

A imagem da figura 8 apresenta a arquitetura da Plataforma de BI *Open Source Pentaho*. A camada de Aplicações de Terceiros representa os dados de diversas fontes. A

camada de Integração dos Dados e da Aplicação é responsável pelo armazenamento dos dados. A camada de *Business Intelligence* representa os modelos dimensionais do negócio. Acima desta camada estão os módulos da Ferramenta *Pentaho* e por fim a camada de Apresentação, responsável pela interação com o usuário final (*Pentaho*, 2010).

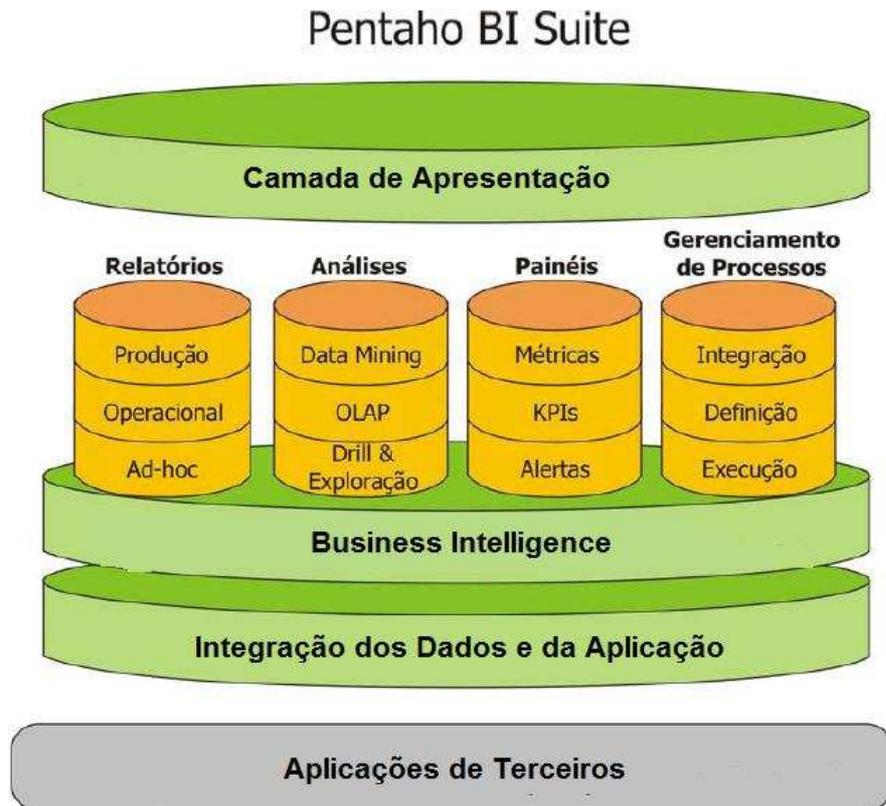


Figura 8: Arquitetura da Plataforma do BI *Pentaho* (adaptação, *Pentaho*, 2010)

3.2 Necessidades da Ferramenta

A Ferramenta de BI *Open Source Pentaho* recomenda uma configuração mínima de hardware e software para que se tenha um bom desempenho dos recursos oferecidos. O computador deve ter no mínimo 2 GB de memória *RAM*, 1 GB de espaço disponível no disco rígido, um processador *AMD64 Dual-core* ou *EM64T* e sistemas operacionais que trabalhem com 32-bit (*Pentaho*, 2010).

Em termos de sistemas operacionais são oficialmente suportados no *Windows XP* com *Service Pack 2*, *Solaris 10*, *Mac OS X 10.4*, as distribuições modernas do *Linux* (*SUSE Linux*

Enterprise Desktop e Server 10 e o Red Hat Enterprise Linux 5 são oficialmente suportados, mas na maioria dos outros deve funcionar). Não importa em qual sistema operacional ela será utilizada, é necessário ter instalado o *Sun Java Runtime Environment (JRE)* versão 1.5 (pode ser referenciado às vezes como a versão 5.0) e as variáveis de ambiente: *JAVA_HOME* e *PATH* estejam devidamente configurados. Nas estações de trabalho é necessário um navegador *Web* instalado com as versões mínimas como: a *Internet Explorer 6*, *Firefox 3.0* (ou o *Mozilla Netscape* ou equivalente) e *Safari 2.0.3* (Pentaho, 2010).

Os arquivos para *download* encontram-se no endereço: <http://sourceforge.net/projects/pentaho/> (Source Forge, 2010), após deve-se descompactar o arquivo em um diretório, entrar na pasta */pentaho/server/biserver-ce/* clicar em *star-pentaho.bat* para *Windows* ou *star-pentaho.sh* para *Linux*, para que seja inicializado o servidor e por fim acessar via navegador *Web* o endereço: <http://localhost:8080/pentaho>.

3.3 Pentaho

Os componentes *open source* que integram a criação de soluções de BI *Pentaho* são: *Pentaho Data Intregation*, *Mondrian*, *JPivot*, *Schema Workbench*, *Analysis View* e *Report Designer*.

3.3.1 Pentaho Data Integration

O *Pentaho Data Integration (PDI)* é também conhecido como *Kettle*. Esta ferramenta é utilizada para fazer o processo de *ETL* explicado na seção 2.3, através de uma interface gráfica, possui a licença *LGPL* e foi desenvolvida toda em *Java*. O *PDI* pode trabalhar diretamente conectado a qualquer banco de dados ou utilizar arquivos (Pentaho, 2010). O processo de *ETL* inicia com a definição da fonte dos dados, das transformações e normalizações necessárias e, por fim, normalmente são criados ou carregados os *DWs* ou *Data Marts*.

O *PDI* pode ser dividido em três aplicativos *Spoon*, *Pan* e *Kitchen*.

Spoon: é uma ferramenta gráfica usada para modelar as *transformations* e os *jobs*. *Transformations* é o modelo que lê o fluxo de dados de entrada e faz uma transformação até

atingir o fluxo de dados de saída e *job* é um modelo o qual é programado para ser executado em intervalos determinados, que tem entradas com as transformações de *downloads*, *FTP*, entre outras e saídas como envio de *e-mails*, por exemplo. Na modelagem usando estes modelos existem dois conceitos importantes: *Step* e *Hop*. *Step* (um passo) é uma execução de uma tarefa específica, como normalizar um conjunto de dados. Os *Steps* são agrupados em categorias como *Transform*, *Joins*, *Input*, *Output*, entre outras. *Hop* é um modelo gráfico de um fluxo de dados entre dois *Steps*, com uma origem e um destino (Pentaho, 2010).

Pan é a ferramenta responsável para a execução das *transformations* criadas (Pentaho, 2010).

Kitchen é responsável pela execução dos modelos *jobs* criados (Pentaho, 2010).

3.3.2 Mondrian/JPivot

Mondrian é um servidor *OLAP*, responsável por converter consultas feitas em linguagem *Multidimensional Expressions (MDX)* em consultas *SQL*. Ele lê os dados de bases relacionais e os mostra no formato multidimensional.

A linguagem *MDX* é uma estrutura criada com a finalidade de fazer consultas em modelos multidimensionais, normalmente em um *DW* ou *Data Mart*. Esta linguagem foi elaborada pela *Microsoft* (Microsoft, 2010).

As consultas possuem a seguinte estrutura básica:

```
SELECT <set> ON COLUMNS,
      <set> ON ROWS
FROM [<nome_cubo>]
```

“*Set*” é definido por um conjunto de membros de uma dimensão ou um cruzamento de várias dimensões (Microsoft, 2010).

Exemplo:

```
SELECT {[Measures].[Consumidor_Energia], [Measures].[Consumidor_VIPago]}
ON COLUMNS
{[Tempo].[2010]} ON ROWS
FROM [ConsumoEnergia]
WHERE ([Cidade].[Farroupilha])
```

O exemplo acima faz uma consulta no Cubo ([ConsumoEnergia]) a cada uma das

dimensões de [Tempo], [Cidade] e alguns dos seus membros. O comando *SELECT* define os eixos da consulta da dimensão Tempo. O comando *FROM* identifica em qual cubo será feita a pesquisa, neste caso a consulta será feita no cubo [ConsumoEnergia]. O comando *WHERE* restringe a consulta a ser realizada em um determinado intervalo de dados, neste caso irá retornar somente os dados referentes à cidade de Farroupilha.

Usado juntamente com o Mondrian, o *JPivot* é um cliente *OLAP*. A ferramenta *JPivot* processa o resultado obtido através do servidor *Mondrian* e ilustra as informações em uma interface gráfica. As visualizações destas informações podem ser ilustradas em diversos formatos, como por exemplo, tabelas, gráficos, entre outras, para o usuário final.

3.3.3 *Schema Workbench*

O *Schema Workbench* é uma interface gráfica, responsável por criar, editar ou testar esquemas que definem os cubos *OLAP*, que serão utilizados e analisados pela ferramenta de Análise do *Pentaho* (*Mondrian/JPivot*). Esta ferramenta cria os esquemas que geram os arquivos *XML*, o qual contém a descrição de quais serão as dimensões, cubos e suas tabelas fatos e dimensões associadas. Além disso, é gerado um *XACTION* a partir desse *XML*, pois é o arquivo que será identificado pelo *Pentaho*. A publicação do *XML* é feita automaticamente por esta ferramenta e ela tem uma integração com o *Pentaho* (*Pentaho*, 2010).

O *Schema Workbench* oferece um ambiente para testar consultas *MDX* dentro de um esquema criado (*Pentaho*, 2010).

3.3.4 *Analysis View*

O módulo *Analysis View* é o responsável por gerar as informações resultantes da análise do cubo. As exibições são semelhantes aos relatórios, com exceção que elas são dinâmicas e interativas, possibilitando ao usuário personalizar a sua consulta para que ele consiga obter um resultado satisfatório com o seu objetivo. Esta ferramenta disponibiliza as operações *OLAP* explicada na seção 2.6, para que o usuário possa detalhar e trabalhar mais a consulta sobre os dados do cubo, permitindo buscas mais aprofundadas e detalhadas (*Pentaho*, 2010).

Estas consultas são feitas através do Servidor *Mondrian* e são gerados resultados gráficos através do *JPivot*. Além disso, a ferramenta disponibiliza a possibilidade de exportar as informações em arquivos no formato de *PDF* e planilhas do *Excel*.

3.3.5 *Report Designer*

O *Report Designer* é a ferramenta para criar e gerar relatórios, desde um mais simples até um mais sofisticado e rico em detalhes com uma aparência profissional. Estes relatórios podem ser criados baseados em modelos já pré-estabelecidos, permitindo o usuário personalizá-lo conforme suas necessidades ou iniciar a geração de um novo a partir do nada (*Pentaho*, 2010).

3.4 Exemplo de Uso

Nesta seção apresenta-se alguns exemplos de utilização das ferramentas que compõe a Plataforma de BI *Open Source Pentaho*. É importante destacar que, a primeira etapa a se fazer é a análise e o estudo juntamente com o responsável pela organização para a definição de qual o objetivo pretendido e o resultado que se espera com a utilização da Ferramenta. Nesta seção são demonstrados apenas exemplos de utilização das ferramentas baseados em informações que já estão pré-estabelecidos no *Pentaho*, para testes e auxílio na compreensão da utilização. No próximo capítulo será feito um estudo de caso verídico com aplicação na área da saúde.

Inicialmente é feita a criação do *Data Mart* ou *DW*, normalização e carregamento dos dados, conforme o objetivo pretendido. Nesta etapa são criados diversos processos para realizar o ETL, através do *Pentaho Data Integration*, até a criação completa do *Data Mart* ou *DW*.

A figura 9 apresenta um processo ETL que copia uma tabela do banco de dados operacional para o banco *Data Mart*. Neste exemplo será criada a tabela **tb_dimensao_clientes** no *Data Mart*, onde serão copiados os dados da tabela clientes da base operacional.

O processo segue da seguinte maneira: *Table Input* para buscar os valores da tabela **clientes** para a carga da tabela **tb_dimensao_clientes** (são definidos quais campos serão

carregados). *Select Values* é responsável pela seleção de valores, alteração de nome dos campos e definição dos tipos de dados. *Table Output* é responsável pela criação da tabela no *Data Mart* e sua população dos dados.



Figura 9: Processo ETL dos dados da tabela clientes para a tabela tb_dimensao_clientes

Após a etapa da criação do *Data Mart* ou *DW*, para que sejam feitas as consultas e análises, é necessário a criação de cubos com as dimensões e medidas. Nesta etapa é utilizada a Ferramentas *Schema Workbench*.

A figura 10 mostra a criação de um cubo para consultas na tabela produto, contendo as dimensões de Tempo e Clientes, mais a medida Quantidade.



Figura 10: Criação do Cubo produto_qtde

Neste momento, como já existem o *Data Mart* criado e carregado, e o cubo criado, é possível fazer análises e consultas sobre estas informações. A figura 11 ilustra um resultado possível, mas conforme a utilização dos recursos do *OLAP* na ferramenta *Analysis View*, é possível obter várias visões de diversos formatos e maneiras.

		Quantidade			
		Tempo			
Cientes	Produtos	● Total anos	● + 2003	● + 2004	● + 2005
+ Todos Clientes	- Total Produtos	105.331	36.439	49.417	19.475
	+ Carros Clássicos	35.552	12.762	16.085	6.705
	+ Motos	12.708	4.031	5.906	2.771
	+ Aviões	11.860	3.833	5.820	2.207
	+ Navios	8.499	2.844	4.309	1.346
	+ Trens	2.818	1.000	1.409	409
	+ Ônibus	11.001	4.056	5.024	1.921
	+ Carros Antigos	22.893	7.913	10.864	4.116

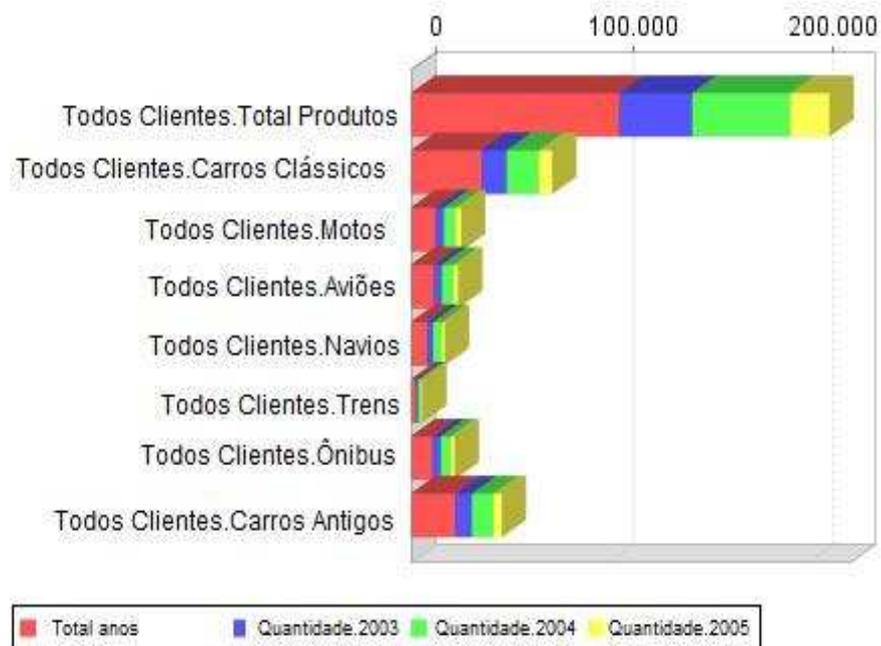


Figura 11: Resultado do cubo produto_qtde através do *Pentaho Analysis* (Pentaho, 2010)

Na figura 11 foi adicionada a dimensão tempo nas colunas, e para as linhas foram adicionadas as dimensões Clientes e Produtos. Após definido o resultado foi criado um gráfico com a finalidade de se ter uma visualização aprimorada e clara das informações.

3.5 Considerações Finais

Este capítulo apresentou o funcionamento e os recursos da ferramenta de BI *Pentaho*, bem como um exemplo de seu uso. No próximo capítulo é explicada quais bases de dados são utilizadas para o estudo de caso e sua origem. Depois, é apresentado o estudo de caso de dois cenários baseados nos dados da área da saúde, explicando ao final as vantagens e desvantagens encontradas.

4 BI NA ÁREA DA SAÚDE

Este capítulo apresenta um estudo de caso da Ferramenta de BI *Pentaho*, através de dois cenários. Os cenários utilizam dados da área da saúde oriundos do DATASUS. Nas seções seguintes são descritas a modelagem multidimensional, a implementação e a visualização de cada cenário. A modelagem é a parte fundamental dos processos de criação de um cubo multidimensional, pois através dela determina-se a visualização das informações para o usuário final. Além da modelagem de cada cenário, descreve-se a criação de *Data Marts*, o uso da ferramenta *Kettle* no processo de ETL e a implementação do cubo pela ferramenta *Schema Workbench*. Ao final do capítulo é feita uma análise do estudo de caso na área da saúde.

4.1 Importância do BI na área da Saúde

O Ministério da Saúde é o órgão do Poder Executivo Federal responsável pelos cuidados, planos e políticas públicas para a assistência e prevenção à saúde da população Brasileira. Ele tem o objetivo de diminuir e controlar as doenças da saúde (Ministério da Saúde, 2005). Dentro do Ministério da Saúde existe um órgão da Secretaria Executiva do Ministério denominado Departamento de Informática do Sistema Único de Saúde (DATASUS). Ele é responsável por coletar, armazenar e distribuir informações sobre a saúde (DATASUS, 2008).

O Brasil com o auxílio dos recursos tecnológicos na área da saúde para armazenar os dados nos sistemas de informações em saúde, gera cada vez mais um volume maior de dados (Ministério da Saúde, 2005). Estes dados são armazenados em diversos bancos de dados relacionais distintos, gerando um grande volume de dados em cada um. Com isso, pode ocorrer a multiplicidade de informações, distribuídas em diversos sistemas, e o preenchimento incompleto dos dados. A tecnologia de BI auxilia na união, integração e padronização dos dados, possibilitando uma melhor visualização e análise das informações. A área da saúde aliada a tecnologia de BI, pode extrair informações do seu banco de dados com uma melhor precisão, exibindo informações úteis e mais precisas. As informações geradas podem mostrar

e auxiliar os órgãos públicos, clínicas, hospitais, entre outras instituições na área da saúde a tomarem decisões mais apropriadas em relação aos cuidados e prevenções necessárias para a melhora da saúde da população brasileira.

4.2 Origem dos Dados

O estudo de caso realizado utiliza dados de dois grandes sistemas de informações do Sistema Único de Saúde (SUS). Eles são o Sistema de Informações sobre Mortalidade (SIM) e o Sistema de Informações de Nascidos Vivos (SINASC).

O SIM possui os dados sobre registros de óbitos de pessoas no território brasileiro através das Declarações de Óbito (DO) coletadas pelas Secretarias Estaduais de Saúde. O DO é um documento padrão nacional, preenchido pelo médico responsável pelo atendimento, sempre que ocorrer um óbito, independente do lugar onde ocorra o óbito. Estas informações podem ser utilizadas por gestores da saúde, pesquisadores ou entidades da sociedade com a finalidade criar programas de prevenção e controle de doenças (DATASUS, 2008). Os campos e estrutura da tabela encontram-se detalhados no Anexo C.

O SINASC possui os dados sobre os nascidos vivos registrados no território brasileiro através da Declaração de Nascimento (DN). Ele possui características importantes como sexo, local onde ocorreu o nascimento, tipo de parto e peso ao nascer, entre outras (DATASUS, 2008). Através destes dados é possível analisar informações sobre a saúde das gestantes e dos recém-nascidos. Os campos e estrutura da tabela encontram-se detalhados no Anexo D.

Os bancos de dados SIM e SINASC utilizados como fonte de dados possuem informações, do ano de 1996 até 2005, para os Estados do Acre, Bahia, Minas Gerais, Rio de Janeiro, Rio Grande do Sul, Pará, Sergipe e Tocantins. É importante salientar que analisando-se os bancos de dados SIM e SINASC identifica-se a falta de preenchimento de algumas informações.

Além dos dois grandes bancos de dados, existe a Tabela da 10^a Revisão Da Classificação Internacional de Doenças (CID-10) que tem o cadastro padronizado e catalogado das doenças e problemas referente à saúde, o qual utiliza o padrão da Nomenclatura Internacional de Doenças (DATASUS, 2008).

4.3 Indicadores na Área da Saúde

O DATASUS considera Indicadores e Dados Básicos para Saúde (IDB) definidos pela Rede Interagencial de Informações para a Saúde (RIPSA). As instituições que compõem a RIPSA são responsáveis pelos principais sistemas de informação de base nacional utilizados: Ministério da Saúde, Instituto Brasileiro de Geografia e Estatística (IBGE), Instituto de Pesquisa Econômica Aplicada (IPEA) e Ministério da Previdência Social. A RIPSA tem o propósito de subsidiar, com informações relevantes, os processos de formulação, gestão e avaliação de políticas e ações públicas de importância estratégica para o sistema de saúde brasileiro (RIPSA, 2008). Os indicadores da RIPSA apresentam o estado de saúde da população. Eles são classificados em seis grupos: A – Indicadores demográficos; B – Indicadores Socioeconômicos; C – Indicadores de Mortalidade; D – Indicadores de Morbidade e fatores de risco; E - Indicadores de Recursos; F - Indicadores de cobertura (RIPSA, 2008).

Os bancos de dados dos sistemas nacionais da área da saúde são distribuídos, e os cadastros das informações são feitos em bancos distintos. Com isso os indicadores podem ser gerados de forma incompleta fazendo com que os gestores na área da saúde tomem decisões equivocadas. Além disso, os dados na maioria das vezes são apresentados em forma de tabelas, dificultando e ocultando informações importantes e úteis para sua interpretação. Mediante estas e outras dificuldades, surgiu o projeto OTICSSS (Observatório de Tecnologias de Informação e Comunicação em Sistemas e Serviços de Saúde).

O projeto OTICSSS tem o foco em analisar e organizar os dados com os recursos tecnológicos, visando desenvolver um cenário que faça a integração dos dados através dos sistemas de informação com aplicações para o ensino e a gestão de saúde. Este projeto é desenvolvido pela Universidade de Caxias do Sul em conjunto com outras instituições de pesquisa no Brasil. Os objetivos deste projeto são: desenvolver e implementar um conjunto de soluções tecnológicas e operacionais, a fim de qualificar o monitoramento e avaliação de indicadores de saúde para quatro regiões (norte, nordeste, sul e sudeste); criar meios que facilitem o acesso e a troca de informações entre os atores envolvidos com informação em saúde, estabelecendo a base para um processo permanente e contínuo de gestão no sistema de saúde (Relatório projeto OTICSSS, apud Sanfelice, 2008).

A finalidade do estudo de caso é utilizar o *Pentaho* como uma ferramenta de análise de

dados oriundas da área da saúde, visando mapear os recursos da ferramenta e analisar os resultados obtidos. O estudo de caso baseou-se em algumas informações e base de dados do Projeto OTICSSS.

4.4 Visão Geral da Implementação

O estudo de caso desenvolvido levou em consideração alguns IDBs do Grupo C – Mortalidade (RIPSA, 2008) para testes e aplicações na integração e análise dos dados.

O Grupo C - Mortalidade é composto por 19 indicadores, sendo que neste trabalho foram implementados 15 deles. A tabela 1 apresenta a lista completa dos indicadores do Grupo C. A segunda coluna indica com um “X” os indicadores que foram implementados. Além disso, as próximas duas colunas identificam se as informações podem ser visualizadas por Estado e também, por Município. A falta de implementações por município se deve a uma incompatibilidade dos dados encontrados entre as tabelas do DATASUS e do IBGE, o que dificultou a tarefa de integração e não pode ser resolvida a tempo neste trabalho.

Tabela 1: Tabela dos indicadores do Grupo C, da RIPSA.

Descrição dos Indicadores do Grupo C-Mortalidade	Imple- mentado	Dimensão Local	
		UF	Município
Taxa de mortalidade infantil - C.1	X	X	X
Taxa de mortalidade neonatal precoce - C.1.1	X	X	X
Taxa de mortalidade neonatal tardia - C.1.2	X	X	X
Taxa de mortalidade pós-neonatal - C.1.3	X	X	X
Taxa de mortalidade perinatal - C.2			
Taxa de mortalidade em menores de 5 anos - C.16	X	X	X
Razão de mortalidade materna - C.3			
Mortalidade proporcional por grupos de causas - C.4			
Mortalidade proporcional por causas mal definidas - C.5	X	X	X
Mortalidade proporcional por doença diarréica aguda em menores de 5 anos - C.6	X	X	X
Mortalidade proporcional por infecção respiratória aguda em menores de 5 anos - C.7	X	X	X
Taxa de mortalidade específica por doenças do aparelho circulatório - C.8	X	X	
Taxa de mortalidade específica por causas externas - C.9	X	X	
Taxa de mortalidade específica por neoplasias malignas - C.10	X	X	
Taxa de mortalidade específica por acidente de trabalho - C.11			
Taxa de mortalidade específica por diabete melito - C.12	X	X	
Taxa de mortalidade específica por aids - C.14	X	X	
Taxa de mortalidade específica por afecções originadas no período perinatal - C.15	X	X	X
Taxa de mortalidade específica por doenças transmissíveis - C.17	X	X	

Cada indicador possui uma ficha de qualificação contendo o que significa o indicador,

qual sua interpretação, sua utilidade e limitação, fonte de dados utilizados, métodos de cálculo e sugestão de categorias de unidades geográficas para análise. Estas informações são encontradas em (IDB, 2010).

A base de dados do DATASUS foi copiada para o banco de dados *Postgresql*, para visualizar e interpretar as tabelas existentes. Foi necessário realizar um estudo sobre os dados que são fornecidos nas tabelas e entender o que significa cada informação, para isso foram utilizados às estruturas de tabelas ilustradas no Anexo C e D.

O indicador de Taxa de Mortalidade Infantil C.1, foi escolhido para iniciar os testes. Foram feitas diversas tentativas de modelagens para tentar obter o resultado satisfatório para o Indicador. Algumas delas foram: criar a tabela fato sumarizada, criar as tabelas de dimensões com os campos da tabela sendo chaves primárias, criado tabelas dimensões com um nível apenas, criadas duas tabelas fatos independentes: uma para os dados da mortalidade e outra para os dados de nascimentos, entre outras. Como não foi obtido sucesso na modelagem do indicador, iniciou-se o processo de uma modelagem simplificada. Ela teve a finalidade de aprender e utilizar as ferramentas, passando por todos processos necessários para se chegar a visualização das informações no *Pentaho*. Ao longo destes processos foi possível aprender e verificar como se comportam as ferramentas e o que é necessário para obter o resultado desejado.

O processo de construção de um DW iniciou-se com a modelagem multidimensional para a criação do *Data Mart*, desenvolvendo as tabelas dimensões e tabelas fatos necessárias. Após foi feito o processo de ETL utilizando a ferramenta *Kettle*, padronizando as informações e integrando os dados. Nesta etapa foi construído e populado o *Data Mart*. Depois dos dados armazenados foi feita a criação do cubo através da ferramenta *Schema Workbench*. Nesta etapa são criados os cubos que são lidos e consultados para a geração das informações no *Pentaho Analysis*. E por fim, é feito o processo de visualização no *Pentaho Analysis*. Nesta última etapa é possível verificar se foi feita a criação do cubo corretamente, pois é a etapa que obtemos o resultado das informações.

Após um período de adaptação ao software, inúmeros testes e pesquisas chegou-se à conclusão que para criar os indicadores seria necessário a criação de três cubos. O primeiro cubo contabiliza os dados referentes a um fato contido no cálculo, exemplo, a quantidade de óbitos. O segundo cubo contabiliza os dados referentes ao segundo fato do cálculo, exemplo, a quantidade de nascidos vivos. E o terceiro cubo realiza a operação descrita na ficha de

qualificação, utilizando-se dos das medidas dos dois cubos anteriores. Além disso, as tabelas dimensões são adicionadas ao *Schema* para que sejam compartilhadas e utilizadas por todos os cubos contidos no esquema, sem a necessidade de repetir as informações. Foram definidas três tabelas dimensões:

- Dimensão Tempo: é composta pelos anos. O período definido foi de 1996 até 2005.
- Dimensão Local: é composto por dois níveis, por Estado e Município de residência.
- Dimensão Causa Básica: é composta por dois níveis. Um nível pelo Grupo de Causas Básicas e outro nível das Causas Básicas definidas na Tabela CID-10.

A tabela 2 apresenta os Estados utilizados na Tabela de Dimensão Local. A tabela contém o código do Estado, a sigla do Estado, a Descrição do Estado e também a descrição do Estado escrita toda em Maiúscula.

Tabela 2: Cadastros de Estados

Código UF	Sigla Estado	Descrição Estado	Descrição Estado em Maiúsculo
52	GO	Goiás	GOIAS
53	DF	Distrito Federal	DISTRITO FEDERAL
0	IG	Ignorado/externo	IGNORADO/EXTERIOR
11	RO	Rondônia	RONDONIA
12	AC	Acre	ACRE
13	AM	Amazonas	AMAZONAS
14	RR	Roraima	RORAIMA
15	PA	Pará	PARA
16	AP	Amapá	AMAPA
17	TO	Tocantins	TOCANTINS
21	MA	Maranhão	MARANHAO
22	PI	Piauí	PIAUI
23	CE	Ceará	CEARA
24	RN	Rio Grande do Norte	RIO GRANDE DO NORTE
25	PB	Paraíba	PARAIBA
26	PE	Pernambuco	PERNAMBUCO
27	AL	Alagoas	ALAGOAS
28	SE	Sergipe	SERGIPE
29	BA	Bahia	BAHIA
31	MG	Minas Gerais	MINAS GERAIS
32	ES	Espírito Santo	ESPIRITO SANTO
33	RJ	Rio de Janeiro	RIO DE JANEIRO
35	SP	São Paulo	SAO PAULO
41	PR	Paraná	PARANA

42	SC	Santa Catarina	SANTA CATARINA
43	RS	Rio Grande do Sul	RIO GRANDE DO SUL
50	MS	Mato Grosso do Sul	MATO GROSSO DO SUL
51	MT	Mato Grosso	MATO GROSSO
99	99	Estado Desconhecido	ESTADO DESCONHECIDO

O desenvolvimento do processo dos indicadores do Grupo C – Mortalidade, assinalados como implementados na Tabela 1, foi o mesmo para todos. Foram selecionados apenas dois indicadores para serem detalhados neste documento na forma de dois cenários. O cenário I trata da consulta ao Grupo de Causas por Doenças do Aparelho Circulatório e o cenário II trata do indicador C.7 – Mortalidade proporcional por infecção respiratória aguda em menores de 5 anos.

4.5 Cenário I: Grupo de Causas por Doenças do Aparelho Circulatório

O cenário I é uma consulta livre para aprendizado e teste do *Pentaho*. Este cenário não corresponde a um indicador significativo para o projeto OTICSSS. O cenário ilustra a quantidade de óbitos por Ano, Estado e Município de residência, dos grupos de causas e causas por mortalidade de doenças do Aparelho Circulatório. A tabela 3 apresenta os grupos de causas por doenças do Aparelho Circulatório que se referem ao código I00 até o I99 da tabela do CID-10.

Tabela 3: Cadastro de Grupos de Causas Básicas por Doenças do Aparelho Circulatório

I00 -I09	Febre reumática aguda e doenças reumáticas crônicas do coração
I10 - I15	Doenças hipertensivas
I20	Doenças isquêmicas do coração
I21	Infarto agudo do miocárdio
I22 -I25	Doenças isquêmicas do coração
I26-I52	Outras doenças cardíacas
I60 - I69	Doenças cerebrovasculares
I70	Arteriosclerose
I71 - I99	Restante do doenças do aparelho circulatório

As Causas por Doenças do Aparelho Circulatório são apresentadas detalhadamente no Anexo G. Os dados utilizados nas tabelas dimensões Tempo e Local, são os mesmos

detalhados na seção anterior.

Nas próximas seções são detalhados os processos para obtenção do resultado final.

4.5.1 Modelagem

A primeira etapa de um processo de construção de um *Data Mart* é a modelagem. Este cenário tem o objetivo de modelar o número de óbitos por doenças do aparelho circulatório, por ano, Estado e Município de residência, por causas e grupo de causas. Foram criadas três tabelas dimensões e uma tabela fato. As tabelas dimensões criadas são: Tempo, Local e Causa Básica; A tabela fato criada é a Mortalidade Causa Básica (figura 12).

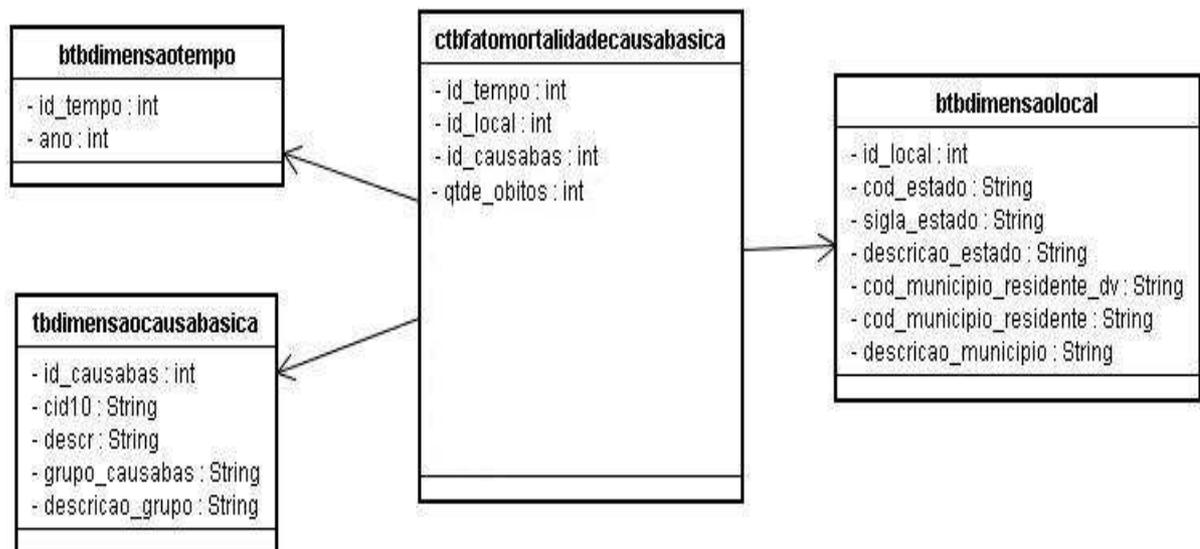


Figura 12: Modelagem multidimensional do Cenário I

A dimensão **btbdimensaotempo** armazena dados referentes ao tempo e o período em que ocorreu o óbito. A Tabela 4 apresenta a estrutura da tabela **btbdimensaotempo**, onde constam os atributos, o tipo de dado, a origem e a descrição. Esta tabela referente a tempo é criada conforme a necessidade de visualização e o formato de que se deseja apresentar os resultados em um determinado período. Esta tabela normalmente tem origem em um arquivo no formato de planilha *Microsoft Excel*.

Tabela 4: Estrutura da Tabela btbdimensaotempo

Atributos	Tipo de Dado	Origem	Descrição
Id_tempo	Inteiro	Criado o campo para ser chave primária.	Chave primária da tabela
Ano	Inteiro	Criado campo para identificar o ano.	Identifica o Ano da ocorrência do óbito.

A dimensão **btbdimensãolocal** armazena dados referentes ao local que o indivíduo que faleceu reside, foi escolhido exibir informações por Estado e Município de residência.

A Tabela 5 apresenta a estrutura da tabela **btbdimensaolocal**, onde constam os atributos, o tipo de dado, origem e a descrição. Nesta tabela é importante ressaltar que existem dois códigos de municípios de residência, devido à diferença de tamanho do tipo de dado. O campo Cod_municipio_residente_dv se refere ao código do município encontrado na tabela SIM, e o campo Cod_municipio_residente se refere ao código do município encontrado na tabela do IBGE. Na tabela foram criados os dois campos de códigos de municípios, com o intuito de poder utilizá-lo na criação dos indicadores do Grupo C, da RIPSA, que utilizam informações referentes à tabela do IBGE.

Tabela 5: Estrutura da Tabela btbdimensaolocal

Atributos	Tipo de Dado	Origem	Descrição
Id_local	Inteiro	Criado o campo para ser chave primária.	Chave primária da tabela
Cod_estado	String	Tabela cadmun campo ufcod	Identifica o código do Estado de residência da pessoa que morreu.
Sigla_estado	String	Tabela caduf campo ufsigla	Sigla do Estado
Descricao_estado	String	Tabela caduf campo ufnomeex	Descrição do Estado

Cod_municipio_residente_dv	String	Tabela cadmun campo mundcoddv	Identifica o código do Município de residência da pessoa que morreu.
Cod_municipio_residente	String	Tabela cadmun campo mundcod	Identifica o código do Município de residência da pessoa que morreu conforme tabela do IBGE.
Descricao_municipio	String	Tabela cadmun campo munnomeex	Nome do município.

A dimensão **tbdimensaocausabasic** armazena dados referentes à causa do óbito. Esta tabela é constituída pelas causas básicas e pelo grupo ao qual pertence a causa. Esta dimensão possibilita a visualização da quantidade de óbitos por causa ocorrida e/ou por grupo de causas.

A tabela 6 apresenta a estrutura da tabela **tbdimensaocausabasic**, onde constam os atributos, o tipo de dado, origem e descrição.

Tabela 6: Estrutura da Tabela tbdimensaocausabasic

Atributos	Tipo de Dado	Origem	Descrição
Id_causabas	Inteiro	Criado o campo para ser chave primária.	Chave primária da tabela
cid10	String	Tabela cid10 campo cid10	Identifica a causa básica do óbito.
Descr	String	Tabela cid10 campo descr	Descrição da Causa Básica.
Grupo_causabas	String	Tabela grupo_causabasic campo código_grupo	Identifica o código do grupo que a causa básica pertence.
Descricao_grupo	String	Tabela grupo_causabasic campo descrição_grupo	Descrição do grupo de Causa Básica

A tabela fato **ctbfatomortalidadecausabasic** é composta pelas chaves primárias das tabelas de dimensão, sendo nesta tabela representadas como chaves estrangeiras. Além disso,

na tabela 6 são adicionados os campos que servem como medidas para a visualização dos dados.

Na tabela SIM, onde buscamos as informações dos óbitos, possui dados descritivos nominais que identifica o óbito. Com isso, foi criado um campo `qtde_obitos` que por padrão recebe o valor um (1) para a identificação da medida em formato numérico, necessária para a modelagem.

A tabela 7 apresenta a estrutura da tabela **ctbfatomortalidadecausabasic**, onde constam os atributos, o tipo de dado, origem e descrição.

Tabela 7: Estrutura da Tabela `ctbfatomortalidadecausabasic`

Atributos	Tipo de Dado	Origem	Descrição
Id_tempo	Inteiro	Tabela <code>btbdimensaotempo</code> campo <code>id_tempo</code>	Identificação do Tempo
Id_local	Inteiro	Tabela <code>btbdimensaolocal</code> campo <code>id_local</code>	Identificação do Local
Id_causabas	Inteiro	Tabela <code>tbdimensaocausabasic</code> campo <code>id_causabas</code>	Identificação da Causa Básica
Qtde_obitos	Inteiro	Criado o campo para servir como medida.	Identificação da quantidade de óbitos.

4.5.2 Processo de *ETL*

Na realização do processo de ETL foi utilizada a Ferramenta *Kettle*. Neste processo são geradas as tabelas dimensões e a tabela fato para a construção do *Data Mart*. Os dados das tabelas do banco de dados operacional são copiados para a criação do *Data Mart*, seguida pela integração dos dados.

Na dimensão `tbdimensaotempo` o processo de ETL criou a tabela através de uma entrada de um arquivo do Excel, o qual contém os anos que serão copiados para a tabela de dimensão. Além disso, foi acrescentado o `id_tempo` para a criação da chave primária e

identificação da tabela de tempo.

A figura 13 ilustra as etapas seguidas para a criação e carga dos dados para a tabela *tbdimensaotempo* que é gerada no *Data Mart*.



Figura 13: ETL para tabela de dimensão tempo

Os componentes utilizados no processo de ETL para a tabela de dimensão tempo estão descritos abaixo, contendo ao lado de cada transformação o nome do componente utilizado pelo Kettle:

1 – Entrada Tempo (componente *Excel Input*): busca no arquivo do *Excel* os dados para carga na tabela *tbdimensaotempo*.

2 – Selecao Valores (componente *Select Values*): seleção dos campos desejados para adicionar a tabela.

3 – Adiciona Id (componente *Add Sequence*): adição de um campo de sequência para criar a chave primária para a dimensão.

4 – Seleciona Valores 2 (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários na composição da Dimensão.

5 – *tbdimensaotempo* (componente *Table Output*): responsável pela criação da tabela de saída e população dos dados para o *Data Mart*.

Na dimensão *btbDimensaoLocal* o processo de ETL é criado através da Tabela de Municípios, após são localizadas informações referentes ao Estado e é adicionado um campo *id_local* para a criação da chave primária e identificação da tabela de local.

A figura 14 ilustra as etapas seguidas para a criação e carga dos dados para a tabela *tbdimensaolocal* que é gerada no *Data Mart*.

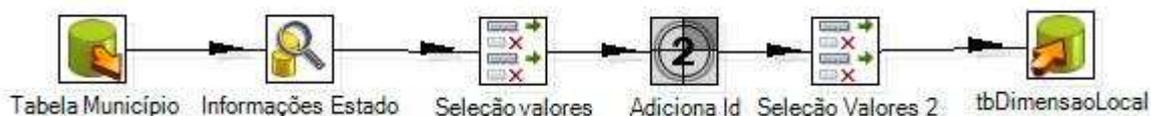


Figura 14: ETL para tabela de dimensão local

Os componentes utilizados para o processo de ETL para a tabela de dimensão local estão descritos abaixo, contendo ao lado de cada transformação o nome do componente utilizado pelo Kettle:

1 – Tabela Município (componente *Table Input*): busca dos valores da tabela município para compor a *tbDimensaoLocal*.

2 – Informações Estado (componente *Database Lookup*): procura informações sobre o estado para acrescentar a *tbDimensaoLocal*.

3 – Seleção valores (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da *tbDimensaoLocal*.

4 – Adiciona Id (componente *Add Sequence*): adição de um campo de sequência para criar a chave primária para a dimensão.

5 – Seleção Valores 2 (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Dimensão.

6 – *tbDimensaoLocal* (componente *Table Output*): responsável pela criação da tabela de saída e população dos dados para o *Data Mart*.

Na dimensão *tbDimensaoCausabasica* foram selecionados somente as causas básicas identificadas na no CID- 10, como doenças do aparelho circulatório. Depois é criada uma fórmula para identificar somente os três primeiros caracteres à esquerda da causa básica para pesquisar à qual grupo de causas esta doença pertence. Além disso, é adicionado um campo *id_causabas* para a criação da chave primária e identificação da tabela de Causa Básica.

A figura 15 ilustra as etapas seguidas para a criação e carga dos dados para a tabela *tbDimensaoCausabasica* que é gerada no *Data Mart*.

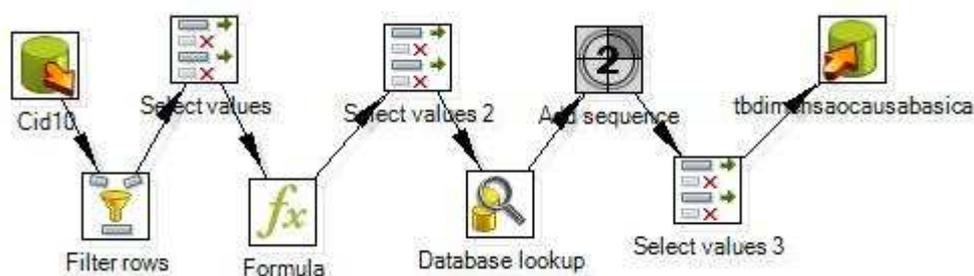


Figura 15: ETL para Tabela de Dimensão Causa Básica

Os componentes utilizados para o processo de ETL para a tabela de dimensão causa básica estão descritos abaixo, contendo ao lado de cada transformação o nome do componente utilizado pelo Kettle:

1 – Cid10 (componente *Table Input*): busca dos valores da tabela Cid10 para compor a tbDimensaoCausabasica.

2 – *Filter rows* (componente *Filter Rows*): criação de um filtro para adicionar somente as causas básicas referentes as doenças do aparelho circulatório.

3 – *Select values* (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Dimensão.

4 – *Formula* (componente *Formula*): criação de uma fórmula para identificação de qual grupo de causa pertence a causa básica em questão. Nesta etapa, a fórmula compara os três primeiros caracteres à esquerda do campo cid10 da Tabela Cid10 com a tabela de Grupo de Causas.

5 - *Select values 2* (componente *Select values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Dimensão.

6 – *Database Lookup* (componente *Database Lookup*): procura informações sobre o grupo de Causas básicas para acrescentar a tbDimensaoLocal.

7 – *Add Sequence* (componente *Add Sequence*): adição de um campo de seqüência para criar a chave primária para a dimensão.

8 – *Select Values 3* (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Dimensão.

9 – tbdimensaocausabasica (componente *Table Output*): responsável pela criação da tabela de saída e população dos dados para o *Data Mart*.

Na tabela Fato *tbFatoMortalidadeCausabasica* o processo de ETL é feito através da carga das informações da tabela SIM, localizando e unindo as chaves primárias das tabelas de dimensões, as quais nesta tabela se tornam chaves estrangeiras. Além disso, é adicionada uma medida *qtde_obitos* com um valor fixo um (1) para contabilizar o número de óbitos.

A figura 16 ilustra as etapas de criação e carga dos dados para a tabela *tbFatoMortalidadeCausabasica* gerada no *Data Mart*.

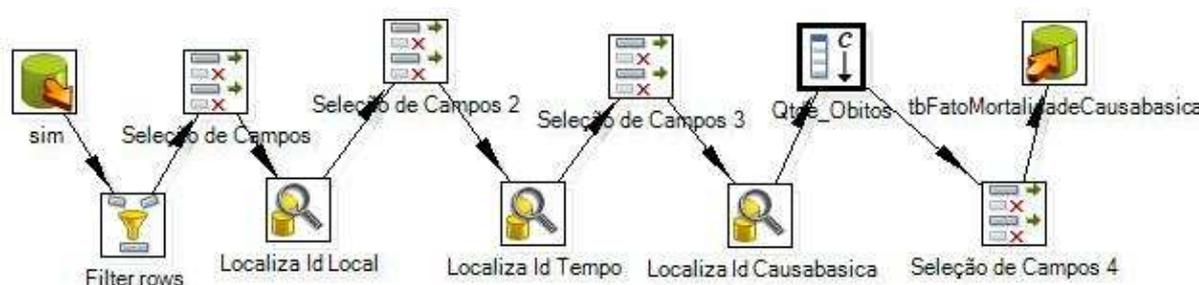


Figura 16: ETL para Tabela Fato de Mortalidade Causa Básica

Os componentes utilizados para o processo de ETL para a tabela de Fato de Mortalidade Causa Básica estão descritos abaixo, contendo ao lado de cada transformação o nome do componente utilizado pelo Kettle:

1 – SIM (componente *Table Input*): busca dos valores da tabela SIM para compor a *tbFatoMortalidadeCausabasica*.

2 – *Filter Rows* (componente *Filter Rows*): criação de um filtro para adicionar somente as causas básicas referentes às doenças do aparelho circulatório.

3 – Seleção de Campos (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

4 – Localiza Id Local (componente *Database Lookup*): localiza o Id Local da tabela dimensão *tbDimensaoLocal* referente ao registro para adicionar a tabela Fato.

5 – Seleção de campos 2 (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

6 – Localiza Id Tempo (componente *Database Lookup*): localiza o Id Tempo da tabela dimensão *tbDimensaoTempo* referente ao registro para adicionar a tabela Fato.

7 – Seleção de Campos 3 (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

8 – Localiza Id Causabásica (componente *Database Lookup*): localiza o Id Causabásica da tabela dimensão tbDimensaoCausabásica referente ao registro para adicionar a tabela Fato.

9 – Qtde_Obitos (componente *Add Constants*): adiciona uma constante qtde_obitos com o valor padrão um (1) para a criação da medida para a Tabela Fato.

7 – Seleção de campos 4 (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

8 – tbFatoMortalidadeCausabásica (componente *Table Output*): responsável pela criação da tabela de saída e população dos dados para o *Data Mart* da tabela Fato.

4.5.3 Implementação

A criação dos cubos é necessária para que a Ferramenta *Pentaho* interprete e faça as consultas na base de dados multidimensionais. As consultas são feitas através da linguagem *MDX*, as quais servem para exibir o resultado final de visualização. Os cubos são criados através de arquivos *XML* denominados de esquema.

Utilizou-se a Ferramenta *Schema Workbench* para a criação de esquemas. É importante salientar nesta etapa que após a criação do esquema é necessário fazer a publicação do mesmo para que a Ferramenta *Pentaho* possa localizá-lo para a realização das consultas.

A figura 17 ilustra a tela com a criação do esquema e cubos utilizados para este cenário. Neste esquema são criadas as dimensões e um cubo com a tabela fato. Nesta situação como é trabalhado apenas com um cubo, pode-se adicionar as dimensões ao *Schema*, figura 17, ou é possível adicionar as dimensões diretamente no cubo. A utilização das dimensões diretamente no esquema tem a finalidade de compartilhar as informações das dimensões para diversos cubos. Ou seja, quando houver dimensões em comum entre os cubos, são acrescentadas estas dimensões ao esquema e cada cubo faz a chamada para as dimensões, sem a necessidade de replicar informações.

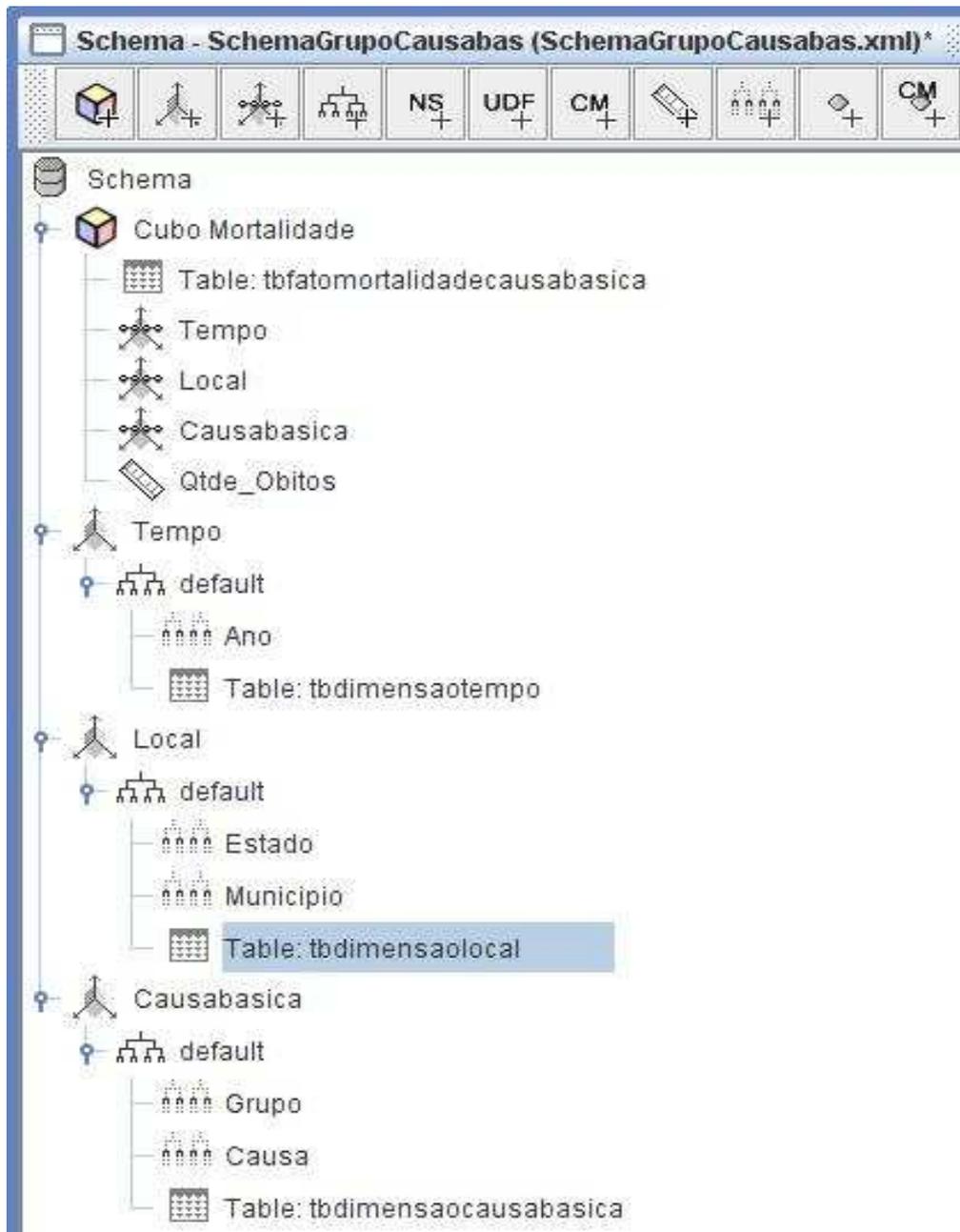


Figura 17: Schema do Cenário I

A figura 17 exibe o cubo Mortalidade com três dimensões: Tempo, Local e Causabásica. Nas dimensões é possível visualizar os níveis e a quantidade de óbitos. Por exemplo, na dimensão Local foram criados dois níveis, um para Estado e outro para Município. Neste caso, a exibição da quantidade de óbitos pode ser por Estado e por Município.

O arquivo XML gerado através do esquema da Figura 18 é apresentado no Anexo E.

4.5.4 Visualização

A Ferramenta *Pentaho Analysis View* foi utilizada para a visualização. Ela é composta pelas ferramentas *Mondrian* e *JPivot*. O *Mondrian* é responsável pela implementação do Servidor *OLAP* e o *JPivot* exibe para o usuário final as consultas *OLAP* realizadas no cubo, através de uma interface gráfica. Nesta interface gráfica é possível fazer consultas dinâmicas, usar filtros e operações conforme aquelas descritas e explicadas na seção 2.6.

A figura 18 ilustra o número de óbitos no ano de 1996 no Estado do RS e no município de Agudo, ocasionados pelos grupos de causas Arteriosclerose e CerebroVasculares. As colunas Tempo, Local e Causabásica são formadas através das tabelas de dimensões. A coluna Qtde_Obitos é a medida incluída na tabela Fato para contabilizar o número de óbitos. Os valores contidos na coluna Qtde_Obitos correspondem ao total de óbitos agrupados conforme as dimensões. O sinal de soma (+) indica que ela pode ser expandida.

Na coluna Tempo, ao clicar-se no sinal de soma (+ All Tempos), observa-se o número de óbitos por ano. Para cada ano é exibido o total de óbitos por Local (+ All Locals), que pode ser estendido por Estado e Município. Para cada Estado ou Município é possível visualizar o número de óbitos totalizados por Grupo de Causa Básica do Aparelho Circulatório e sua doença (+All Causabásicas).

Algumas das informações que podem ser visualizadas na figura 18 são as seguintes:

- O número total de óbitos do ano de 1996 ocorridos no RS no município de Agudo causados pela doença I63.9 Infarto cerebral NE (pertencente ao grupo de causas básicas Doenças cerebrovasculares) é 2.

- O número total de óbitos do ano de 1996 ocorridos no RJ, causados por doenças do aparelho circulatório é 35.812.

- O número total de óbitos do ano de 1996 até 2005 causados por doenças do aparelho circulatório é 970.436.

The screenshot shows the Pentaho Analysis View interface. The main table displays a hierarchy of data for the measure 'Qtde_Obitos'. The hierarchy is as follows:

- Tempo: All Tempos (970.436)
 - Local: All Locals (92.806)
 - AC (376)
 - MG (28.806)
 - PA (3.410)
 - RJ (35.812)
 - RS (22.277)
 - AGUA SANTA (3)
 - AGUDO (36)
 - Aterosclerose (1)
 - I70.9 Aterosclerose generalizada e a NE (1)
 - Doencas cerebrovasculares (9)
 - I60.9 Hemorragia subaracnoide NE (1)
 - I63.9 Infarto cerebral NE (2)
 - I64 Acid vasc cerebr NE como hemorrag isquemico (6)
 - Doencas hipertensivas (1)
 - Doencas isquemicas do coracao (1)
 - Infarto agudo do miocardio (15)

Figura 18: Resultado exibido na ferramenta *Pentaho Analysis View*

4.6 Cenário II: Indicador C.7 - Mortalidade Proporcional por Infecção Respiratória Aguda em Menores de 5 Anos

O cenário II exibe o percentual de óbitos por infecção respiratória aguda em relação ao total de óbitos de menores de cinco anos, por ano, Estado e Município de residência.

4.6.1 Modelagem

A modelagem do cenário II contempla o indicador da área da Saúde C.7 (Mortalidade Proporcional por Infecção Respiratória aguda em menores de 5 anos), do Grupo C de Mortalidade (RIPSA, 2008). A tabela SIM foi utilizada como fonte de dados para a criação do modelo (Anexo C).

O indicador C.7 utiliza o seguinte cálculo: (Número de óbitos de residentes menores

de cinco anos por infecção respiratória aguda / Número total de óbitos de residentes menores de cinco anos por causas definidas) * 100. Os óbitos por infecção respiratória aguda correspondem aos códigos J00 a J22 do capítulo X –Doenças do aparelho respiratório, da CID- 10 (RIPSA, 2008).

No modelo é possível visualizar o indicador por ano, por Estado e Municípios de residência. A partir das análises e detalhes informados foram criadas duas tabelas dimensões e duas tabelas fatos. As tabelas dimensões criadas são: Tempo e Local. As tabelas fatos criadas são: Tabela Fato Mortalidade Menores de cinco anos por doenças de infecção respiratória Aguda e a Tabela Fato Mortalidade Menores de cinco anos. A criação de duas tabelas fatos é necessária para fazer o cálculo no momento da criação do cubo por que os dados em cada tabela serão respectivamente o numerador e o denominador na fórmula de cálculo do indicador. Estas tabelas (dimensões e fatos) são necessárias para o armazenamento dos dados e a modelagem multidimensional utilizada pela Ferramenta de *BI Pentaho*.

A figura 19 ilustra a modelagem multidimensional descrita.

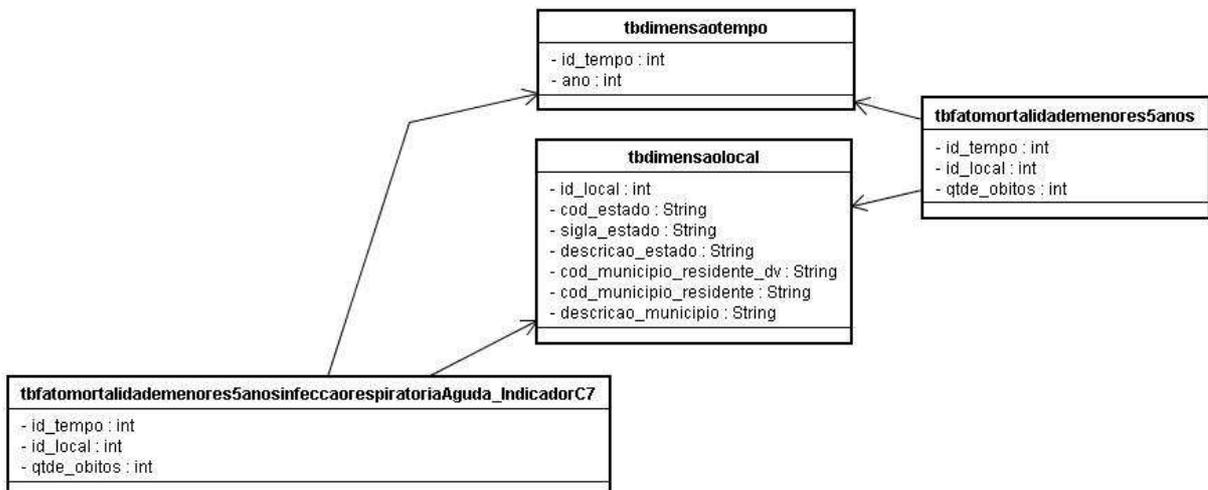


Figura 19: Modelagem Multidimensional do Cenário II

No modelo o usuário pode-se visualizar o resultado referente ao Indicador C.7 por ano, Estado e Município de residência.

As dimensões **tbdimensaotempo** e **tbdimensaolocal** armazenam dados referentes ao tempo e localização do óbito ocorrido. Estas tabelas são descritas e apresentadas na seção 4.3.1.

A tabela fato **tbfatomortalidademenores5anosinfeccaorespiratoriaAguda_IndicadorC7** é composta pelas chaves primárias das tabelas de dimensão, onde nesta tabela elas são representadas como chaves estrangeiras. Além disso, nesta tabela são adicionados os campos que servem como medidas para a visualização dos dados.

Na tabela SIM, onde busca-se as informações dos óbitos, não existem medidas pois cada registro identifica um óbito. Com isso, foi criado um campo qtde_obitos que por padrão recebe o valor um (1) para a identificação da medida.

A tabela fato é composta pelos registros dos óbitos com idade menor ou igual a cinco anos e ocorridos pelas causas de infecção respiratória aguda.

A tabela 8 apresenta a estrutura da tabela **tbfatomortalidademenores5anosinfeccaorespiratoriaAguda_IndicadorC7**, onde constam os atributos, o tipo de dado, origem e descrição.

**Tabela 8: Estrutura da Tabela
tbfatomortalidademenores5anosinfeccaorespiratoriaAguda_IndicadorC7**

Atributos	Tipo de Dado	Origem	Descrição
Id_local	Inteiro	Tabela tbdimensaotempo campo id_tempo	Identificação do Tempo
Id_local	Inteiro	Tabela tbdimensaolocal campo id_local	Identificação do Local
Qtde_obitos	Inteiro	Criado o campo para servir como medida.	Identificação da quantidade de óbitos.

A tabela fato **tbfatomortalidademenores5anos** é composta pelas chaves primárias das tabelas de dimensão, onde nesta tabela elas são representadas como chaves estrangeiras. Além disso, nesta tabela são adicionados os campos que servem como medidas para a visualização dos dados.

Na tabela SIM, onde buscamos as informações dos óbitos, não existem medidas pois cada registro identifica um óbito. Com isso, foi criado um campo qtde_obitos que por padrão recebe o valor um (1) para a identificação da medida.

A tabela fato é composta pelos registros dos óbitos com idade menor ou igual a cinco

anos.

A tabela 9 apresenta a estrutura da tabela **tbfatomortalidademenores5anos**, onde constam os atributos, o tipo de dado, origem e descrição.

Tabela 9: Estrutura da Tabela **tbfatomortalidademenores5anos**

Atributos	Tipo de Dado	Origem	Descrição
Id_local	Inteiro	Tabela tbdimensaotempo campo id_tempo	Identificação do Tempo
Id_local	Inteiro	Tabela tbdimensaolocal campo id_local	Identificação do Local
Qtde_obitos	Inteiro	Criado o campo para servir como medida.	Identificação da quantidade de óbitos.

4.6.2 Processo de ETL

Neste processo são criadas as tabelas dimensões e as tabelas fatos para a construção do *Data Mart*. Neste momento são copiados os dados das tabelas do banco de dados operacional para a criação do *Data Mart* e também é feita a integração dos dados.

Nas dimensões tbdimensaotempo e tbdimensaolocal o processo de ETL é o mesmo explicado na seção 4.3.2.

Na tabela Fato **tbfatomortalidademenores5anosinfeccaorespiratoriaAguda_IndicadorC7** o processo de ETL é feito através da carga das informações da tabela SIM, localizando e unindo as chaves primárias das tabelas de dimensões, as quais nesta tabela se tornam chaves estrangeiras. Além disso, é adicionada uma medida qtde_obitos com um valor fixo um (1) para ser possível contabilizar o número de óbitos e filtrando somente os registros com idade menor ou igual a cinco com causa pertencente a infecção respiratória Aguda.

A figura 20 ilustra as etapas feitas para a criação e carga dos dados para a tabela **tbfatomortalidademenores5anosinfeccaorespiratoriaAguda_IndicadorC7** que é gerada no *Data Mart*.



Figura 20: ETL para Tabela Fato de Mortalidade Menores de 5 anos por Infecção Respiratória Aguda

As etapas e componentes da ferramenta utilizadas para executar o processo são:

1 – Sim (componente *Table Input*): busca dos valores da tabela SIM para compor a `tbfatomortalidademenores5anosinfeccaorespiratoriaAguda_IndicadorC7`.

2 – Filtra Informações (componente *Filter Rows*): criação de um filtro para adicionar somente os óbitos com idade menores de 5 anos referentes às doenças por Infecção Respiratória Aguda.

3 – Seleção de campos (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

4 – Localiza Id Local (componente *Database Lookup*): localiza o Id Local da tabela dimensão `tbDimensaoLocal` referente ao registro para adicionar a tabela Fato.

5 – Seleção de Campos 2 (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

6 – Localiza Id Tempo (componente *Database Lookup*): localiza o Id Tempo da tabela dimensão `tbDimensaoTempo` referente ao registro para adicionar à tabela Fato.

7 – Seleção de Campos 3 (componente *Select Value*s): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

8 – Qtde_obitos (componente *Add Constants*): adiciona uma constante `qtde_obitos` com o valor padrão um (1) para a criação da medida para a Tabela Fato.

9 – Seleção de campos 4 (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

10 – `tbFatoMortalidadeMenores5anosInfeccaoRespiratoriaAgudaIndicadorC7` (componente *Table Output*): responsável pela criação da tabela de saída e população dos dados para o *Data Mart* da tabela Fato.

Na tabela Fato **tbfatomortalidademenores5anos** o processo de *ETL* é feito através da carga das informações da tabela SIM, localizando e unindo as chaves primárias das tabelas de dimensões, as quais nesta tabela se tornam chaves estrangeiras. Além disso, é adicionada uma medida *qtde_obitos*, com um valor fixo um (1), para ser possível contabilizar o número de óbitos. Além disso, é feito um filtro para carregar somente os registros com idade menor ou igual a cinco.

A figura 21 ilustra as etapas feitas para a criação e carga dos dados para a tabela **tbfatomortalidademenores5anos** que é gerada no *Data Mart*.



Figura 21: ETL para Tabela Fato de Mortalidade Menores de 5 anos

As etapas e componente utilizados para a realização do processo apresentado na figura 21 são:

1 – Sim (componente *Table Input*): busca dos valores da tabela SIM para compor a **tbfatomortalidademenores5anos**.

2 – Filtra Informações (componente *Filter Rows*): criação de um filtro para adicionar somente os óbitos com idade menor de 5 anos.

3 – Seleção de Campos (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

4 – Localiza Id Local (componente *Database Lookup*): localiza o Id Local da tabela dimensão **tbDimensaoLocal** referente ao registro para adicionar a tabela Fato.

5 – Seleção de Campos 2 (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

6 – Localiza Id Tempo (componente *Database Lookup*): localiza o Id Tempo da tabela dimensão **tbDimensaoTempo** referente ao registro para adicionar a tabela Fato.

7 – Seleção de Campos 3 (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

8 – Qtde_obitos (componente *Add Constants*): adiciona uma constante qtde_obitos com o valor padrão um (1) para a criação da medida para a Tabela Fato.

9 – Seleção de Campos 4 (componente *Select Values*): seleção dos valores, tipos de dados e alteração de nomes para os campos necessários para a composição da Tabela Fato.

10 – tbFatoMortalidadeMenores5Anos (componente *Table Output*): responsável pela criação da tabela de saída e população dos dados para o *Data Mart* da tabela Fato.

4.6.3 Implementação

A figura 22 apresenta a tela com a criação do esquema e cubos utilizados para este cenário. Neste esquema são adicionadas as dimensões ao esquema e criado três cubos. No primeiro cubo foi adicionado a tabela fato `tbfatomortalidademenores5anosinfeccaorespiratoriaAguda_IndicadorC7`, no segundo cubo foi adicionado a tabela fato `tbfatomortalidademenores5anos`. O terceiro cubo, denominado cubo virtual, tem o objetivo de executar o cálculo referente ao Indicador C.7. As tabelas dimensões adicionadas são Tempo e Local. Neste caso, foram adicionadas as dimensões ao esquema, pois as mesmas são compartilhadas com todos os cubos.

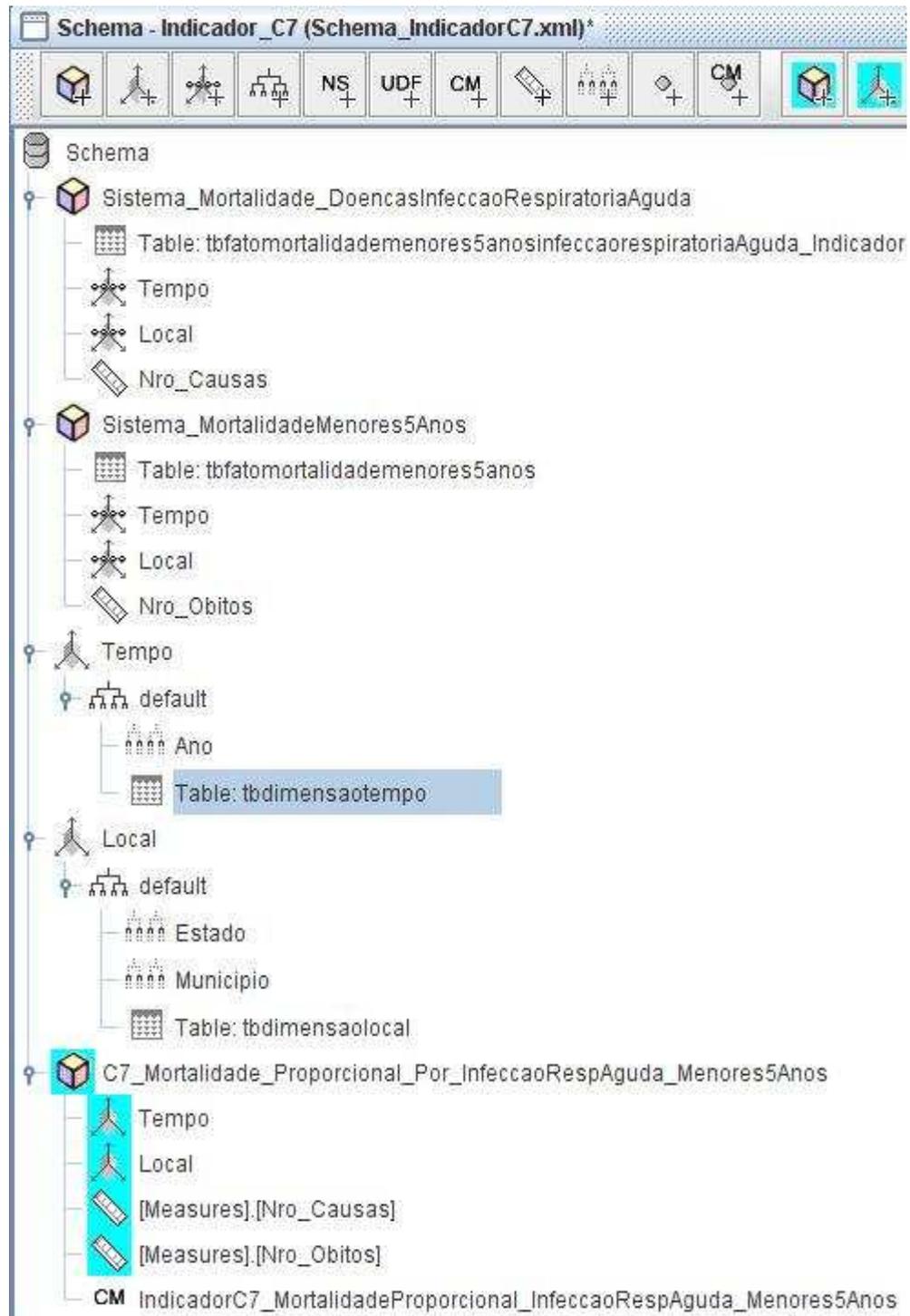


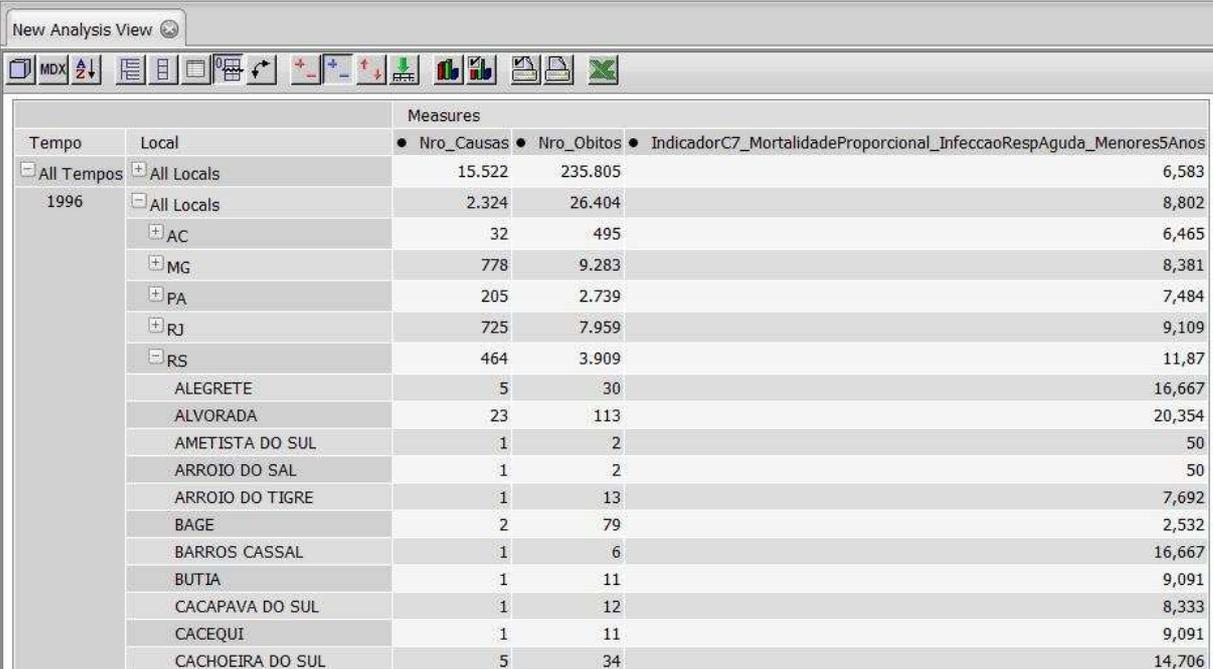
Figura 22: Schema do Cenário II

O arquivo XML gerado através do esquema da Figura 23 é apresentado no Anexo F.

4.6.4 Visualização

A visualização é feita através da Ferramenta *Pentaho Analysis View*. A figura 23 exibe o número de óbitos no ano de 1996 no Estado do RS, no município de Agudo, pelo grupo de Causa Arteriosclerose e CerebroVasculares, com suas respectivas causas.

A figura 23 apresenta nas colunas Tempo e Local, o ano, Estado e Município selecionados nas tabelas de dimensões. A coluna Nro_Causas mostra o número de óbitos com idade inferior a 5 anos causadas por infecção respiratória aguda. A coluna Nro_Obitos apresenta o número de óbitos com idade inferior a 5 anos. E a coluna IndicadorC7_MortalidadeProporcional_ InfeccaoRespAguda_Menores5Anos apresenta o indicador de Mortalidade Proporcional por Infecção Respiratória Aguda em Menores de 5 anos – C.7.



Tempo	Local	Measures		
		• Nro_Causas	• Nro_Obitos	• IndicadorC7_MortalidadeProporcional_ InfeccaoRespAguda_Menores5Anos
All Tempos	All Locals	15.522	235.805	6,583
1996	All Locals	2.324	26.404	8,802
	AC	32	495	6,465
	MG	778	9.283	8,381
	PA	205	2.739	7,484
	RJ	725	7.959	9,109
	RS	464	3.909	11,87
	ALEGRETE	5	30	16,667
	ALVORADA	23	113	20,354
	AMETISTA DO SUL	1	2	50
	ARROIO DO SAL	1	2	50
	ARROIO DO TIGRE	1	13	7,692
	BAGE	2	79	2,532
	BARROS CASSAL	1	6	16,667
	BUTIA	1	11	9,091
	CACAPAVA DO SUL	1	12	8,333
	CACEQUI	1	11	9,091
	CACHOEIRA DO SUL	5	34	14,706

Figura 23: Resultado exibido na Ferramenta *Pentaho Analysis View*

4.7 Considerações Finais

Os estudos de caso desenvolvidos foram muito importantes para avaliar a ferramenta

e o comportamento dos dados. Baseado em dois cenários, os estudos de caso mostraram os passos, etapas e processos utilizados, além dos resultados encontrados. Foram feitas inúmeras tentativas e modelagens na ferramenta de BI *Pentaho*, para se chegar a um resultado satisfatório.

O *Pentaho* utiliza um conjunto de ferramentas para exibir e visualizar os resultados, para isso foi necessário aprender e configurar todas elas, para fazer o processo completo. Apesar de não conhecer as ferramentas, a tecnologia de BI e nem os dados, a maior dificuldade encontrada foi a criação e estruturação da modelagem dos dados, o que demanda tempo e auxílio de um especialista na área para que seja executado com sucesso. Esta dificuldade ocorre devido à complexidade da modelagem e a necessidade da compreensão do assunto que será tratado e exibido.

A ferramenta utiliza uma modelagem multidimensional baseada em Tabelas Fatos, que como o próprio nome diz, são fatos e questões que ao visualizar o resultado final é possível identificar expectativas, tendências e riscos, possibilitando o trabalho com os dados na área da saúde.

Os dados trabalhados possuem um enorme volume de informações, que demanda tempo na hora de fazer o processo de ETL. Nesta fase são criadas as consistências para o banco de dados e no momento que está carregando, caso ocorra algum erro o processo para e a ferramenta exibe um log de informações explicando o motivo do erro. Com isso é necessário ajustar o erro e após reiniciar o processo, demandando tempo. Devido às inconsistências de informações no banco de dados foi necessário criar um cadastro para o município (99 – Município Ignorado) na tabela município, a fim de continuar o processo normalmente. Após todos os processos bem elaborados e trabalhados, o resultado foi eficiente.

Estes cenários apresentados são apenas uma pequena demonstração do que a ferramenta *Pentaho* e a tecnologia de BI podem suportar e fazer, as ferramentas utilizadas são bem completas e amplas, devido à falta de tempo não foi possível aprofundar-se mais em cada ferramenta.

5 CONSIDERAÇÕES FINAIS

Este capítulo apresenta as conclusões obtidas no trabalho elaborado e estudo de caso demonstrado. Por fim, são sugeridos alguns trabalhos futuros identificados e que poderão agregar mais conhecimentos sobre a ferramenta de BI *Pentaho*.

5.1 Conclusão

O presente trabalho teve como objetivo verificar o comportamento da tecnologia de BI em relação aos dados da área da saúde no Brasil. Uma tecnologia que possibilita os responsáveis pela área de saúde analisar as informações de um modo mais interativo, dinâmico e de fácil compreensão.

No desenvolvimento do trabalho se fez uso da ferramenta de BI *Pentaho*, por sua eficiência e dinâmica na demonstração dos resultados, e porque possibilita integrar um grande volume de dados, mostrando as informações de maneira mais fácil, legível e compreensível. Foram necessários alguns estudos sobre a tecnologia de BI, com o intuito de obter maior conhecimento sobre o assunto e sua importância. Foi realizada uma análise sobre os dados, estrutura e informações dos bancos de dados da área da saúde SIM e SINASC, que serviram como fonte de dados para o estudo de caso. Foi levado em conta os indicadores do Grupo C, da RIPSA, para o desenvolvimento de um *Data Warehouse* da saúde.

Os cenários apresentados para demonstrar o comportamento da ferramenta foram o Grupo de Causas por Doenças do Aparelho Circulatório e Indicador C.7 – Mortalidade Proporcional por Infecção Respiratória Aguda em Menores de 5 Anos. Com isso, foi possível verificar que o *Pentaho* é uma ferramenta viável para trabalhar com os dados da área da saúde. Demonstrou-se que a tecnologia de BI não se aplica somente nas organizações, que buscam ser competitivas, mas para todas que buscam auxílio nas tomadas de decisões em diversas situações, neste caso, na área da saúde.

Uma das contribuições deste trabalho foi o desenvolvimento de uma modelagem multidimensional para os indicadores do Grupo C da RIPSA, criando a infraestrutura necessária para se ter um modelo padrão. Outra contribuição foi a verificação de que o

processo de ETL é muito lento, devido ao grande volume de dados que estão distribuídos entre os bancos de dados SIM e SINASC utilizados na integração, necessária para poder fazer a mineração de dados e buscar padrões utilizados na modelagem criada.

Ao final deste trabalho constatou-se que apesar de atender as necessidades da área da saúde, no Brasil, o baixo número de publicações relacionadas a este assunto indica que o BI é ainda pouco explorado na área da saúde no país.

Além das contribuições citadas, este trabalho possibilitou adquirir um conhecimento sobre a tecnologia de BI, sobre a ferramenta *Pentaho*, modelagem e integração de uma grande quantidade de dados, além das informações relacionadas a área da saúde no Brasil.

5.2 Trabalhos Futuros

Tendo em vista os estudos e a viabilidade da tecnologia de BI, identifica-se a necessidade de um estudo para o desenvolvimento dos indicadores, do Grupo C de Mortalidade, da RIPSA, que não foram implementados neste trabalho. Estes indicadores são: C.2 – Taxa de mortalidade perinatal, C.3 – Razão de mortalidade materna, C.4 – Mortalidade proporcional por grupo de causas e C.11 – Taxa de mortalidade específica por acidente de trabalho. Além disso, ampliar o desenvolvimento da implementação dos indicadores que estão apenas por Estado para Município também (C.8, C.9, C.10, C.12, C.14, C.17). E em alguns indicadores são apresentados na fórmula para ajustar o total da população ao meio do ano, para obter o resultado é necessário elaborar um cálculo, pois atualmente não existe esta informação na tabela. Neste trabalho foi utilizado o total da população do ano todo, então a sugestão é fazer este cálculo para os indicadores que precisam desta informação.

Outra sugestão é um estudo e análise sobre a visualização e geração de relatórios da ferramenta de BI *Pentaho* em um nível detalhado, proporcionando ao usuário as informações de forma desejada.

6 REFERÊNCIAS

BARBIERI, Carlos. **BI – Business Intelligence: Modelagem & Tecnologia**. 1. ed. Rio de Janeiro: *Axcel Books*, 2001.

DATASUS. Departamento de Informática do SUS. Disponível em: <http://www.datasus.gov.br>. Acessado em: Maio de 2010.

DAYAL, Umeshwar, et al. **Data Integration Flows for Business Intelligence**. ACM. Março, 2009.

DELMATER, Rhonda, et. al. **Data Mining Explained: A Manager's Guide to Customer-Centric Business Intelligence**. Boston: *Digital Press*, 2001.

FAYYAD, Usama, et. al. **The KDD Process for Extracting Useful Knowledge from Volumes of Data**. In: *Communications of the ACM*, PP. 27-34, Nov. 1996.

GARTNER. *Gartner Group*. Disponível online em <http://www.gartner.com>, acessado em Fevereiro 2010.

IDB. Indicadores e Dados Básicos. Disponível em: <http://www.datasus.gov.br/idb>. Acessado em: Maio de 2010.

INMON, W. H. **Como Construir o Data Warehouse**. 2. ed. Rio de Janeiro: Campus, 1997.

JASPERSOFT. *Jaspersoft Corporation*. Disponível em: <http://www.jaspersoft.com>. Acessado em: Março de 2010.

KIMBALL, Ralph. **Data Warehouse Toolkit: Técnicas para Construção de Data Warehouses Dimensionais**. São Paulo: Makron Books, 1998.

MACHADO, Felipe Nery Rodrigues. **Projeto de Data Warehouse: Uma visão Multidimensional**. São Paulo: Érica, 2000.

MICROSOFT CORPORATION. *Microsoft Developer Network*. Disponível em: <http://www.msdn.com>. Acessado em: Março de 2010.

MITCHELL, Tom. **Machine Learning**. New York: McGraw-Hill, 1997.

PENTAHO CORPORATION. *Pentaho Open Source Business Intelligence*. Disponível em:

<http://www.pentaho.com>. Acessado em: Fevereiro de 2010.

POSTGRESQL. PostgreSQL Brasil. Disponível em: <http://www.postgresql.org.br>. Acessado em: Março de 2010.

RIPSA. Rede Interagencial de Informações para a Saúde. Disponível em: <http://www.ripsa.org.br>. Acessado em: Maio de 2010.

SANFELICE, S. **Visualização de Dados da Área da Saúde para Apoio à Decisão**. Caxias do Sul, Brasil. Dezembro 2008.

SERRA, Laércio. **A essência do *Business Intelligence***. São Paulo: Berkeley Brasil, 2002.

SOURCEFORGE. *SourceForge*. Disponível em: <http://sourceforge.net>. Acessado em: Março de 2010.

SPAGOWORLD. *Engineering Ingegneria Informatica S.P.A.* Disponível em: <http://www.spagobi.org>. Acessado em: Março de 2010.

THOMSEN, ERIK. **OLAP**: Construindo sistemas de informações multidimensionais. Tradução: Daniel Vieira, Rio de Janeiro: Campus, 2002.

TURBAN, Efrain, et al. **Business Intelligence**: Um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman, 2009.

WALTZ, David, et. al. **Data Mining: A Long-Term Dream**. *IEEE Intelligent Systems*, Vol 14, Nº 6, Junho, 1999.

WEBBER, Carine G., et al. **Relatório do Projeto OTICSSS**. (novembro, 2009), UCS.

ANEXO A – Barra de Ferramentas do Schema Workbench

A figura 24 ilustra os componentes da barra de ferramentas do programa *Schema Workbench*, cada componente está identificado por um número para servir de referência, descrevendo as suas funcionalidades a seguir (Pentaho, 2010).

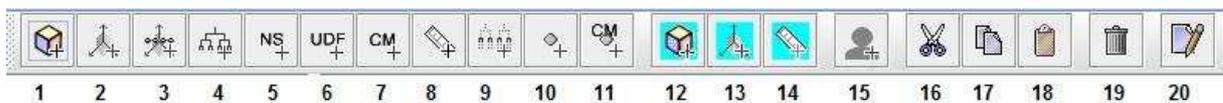


Figura 24: Barra de ferramentas do *Schema Workbench* (Workbench, 2010)

- 1 - *Add Cube*: cria um novo cubo;
- 2 - *Add Dimension*: adiciona uma nova dimensão ao cubo;
- 3 - *Add Dimension Usage*: uso de uma dimensão compartilhada;
- 4 - *Add Hierchy*: adiciona uma nova hierarquia a uma dimensão;
- 5 - *Add Named Set*: define uma fórmula;
- 6 - *Add User Defined Function*: adiciona uma função Java;
- 7 - *Add Calculated Member*: define um novo tipo de medida;
- 8 - *Add Measure*: adiciona uma nova medida ao cubo;
- 9 - *Add Level*: adiciona um novo nível para a hierarquia;
- 10 - *Add Property*: adiciona uma nova propriedade para um nível;
- 11 - *Add Calculated Member Property*: adiciona uma nova propriedade a uma medida;
- 12 - *Add Virtual Cube*: cria um novo cubo virtual;
- 13 - *Add Virtual Cube Dimension*: adiciona uma nova dimensão ao cubo virtual;
- 14 - *Add Virtual Cube Measure*: adiciona uma nova medida ao cubo virtual;
- 15 - *Add Role*: adiciona uma função ao cubo;
- 16 - *Cut*: recorta um item selecionado;
- 17 - *Copy*: copia um item selecionado;
- 18 - *Paste*: cola um item selecionado;
- 19 - *Delete*: delete o item selecionado;
- 20 - *Edit Mode*: exibe o XML criado.

ANEXO B – Barra de Ferramenta do *Analysis View*

A figura 25 ilustra os componentes da barra de ferramenta do *Analysis View* para utilização dos recursos OLAP, cada componente está identificado por um número para servir de referência, descrevendo as suas funcionalidades a seguir (Pentaho, 2010).

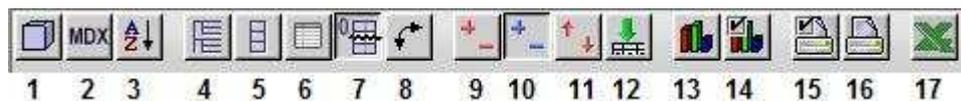


Figura 25: Barra de ferramentas do *Analysis View* (Pentaho, 2010)

1 - *Open OLAP Navigator*: Exibição das informações do cubo, neste momento é possível alterar as estruturas das consultas OLAP, visualizando, removendo ou adicionando dimensões ao cubo. Pode-se alterar a estrutura do cubo, permitindo visualizações de combinações diferentes das dimensões;

2 - *Show MDX Editor*: é responsável por modificar as estruturas do cubo como as funções do comando anterior, porém estas alterações são feitas através da Linguagem MDX;

3 - *Config OLAP Table*: configuração da tabela gerada pelo cubo, ordenação das consultas;

4 - *Show Parent Members*: exibe sim ou não os membros pais (membros superiores na hierarquia);

5 - *Hide Spans*: oculta registros vazios;

6 - *Show Properties*: torna visível as propriedades das dimensões expandidas;

7 - *Suppress Empty Rows/Columns*: suprime colunas ou linhas vazias;

8 - *Swap Axes*: Troca linha por coluna ou vice-versa;

9 - *Drill Member*: exibe todas as ocorrências daquele item;

10 - *Drill Position*: exibe apenas a ocorrência selecionada do item selecionado na árvore;

11 - *Drill Replace*: permite agregar colunas ou linhas durante o Drill Down;

12 - *Drill Through*: explica de onde foi que aquele dado selecionado foi calculado;

13 - *Show Chart*: exibe o gráfico baseado na tabela para o usuário;

14 - *Chart Config*: exibe as configurações que podem ser feitas no gráfico (alterações

de tipo de gráfico, largura, altura , linhas, colunas e outras opções);

15 - *Configure Print Settings*: configuração da impressora;

16 - *Print this Page via PDF*: imprime ou salva arquivos em formato PDF;

17 - *Start Excel*: gera um arquivo no formato XLS.

**ANEXO C – Descrição da estrutura interna da tabela para o SIM
(DATASUS, 2008)**

CAMPO FORMULÁRIO	SEQ	NOME	TIPO/TAM	DESCRIÇÃO
Declaração	1	NUMERODO	C(08)	Número da DO, seqüencial por UF informante e por ano.
7	2	TIPOBITO	C(01)	1: óbito fetal 2: óbito não fetal
8	3	DTOBITO	C(08)	Data do óbito, no formato ddmmaaaa .
8	4	HORAOBITO	C(05)	Hora do falecimento.
10	5	NATURAL	C(03)	Naturalidade, conforme a tabela de países. Se for brasileiro, porém, o primeiro dígito contém 8 e os demais o código da UF de naturalidade.
14	6	DTNASC	C(08)	Data de nascimento do falecido.
15	7	IDADE	C(03)	Idade, composto de dois subcampos. O primeiro, de 1 dígito, indica a unidade da idade, conforme a tabela a seguir. O segundo, de dois dígitos, indica a quantidade de unidades: 0: Idade ignorada, o segundo subcampo e 1: Horas, o segundo subcampo varia de 01 a 23 2: Dias, o segundo subcampo varia de 01 a 29 3: Meses, o segundo subcampo varia de 01 a 11 4: Anos, o segundo subcampo varia de 00 a 99 5: Anos (mais de 100 anos), o segundo subcampo varia de 0 a 99, exemplos: 000: Idade ignorada 020: 20 minutos 103: 3 horas 204: 4 dias 305: 5 meses 400: menor de 1 ano, mas não se sabe o numero de horas, dias ou meses 410: 10 anos 505: 105 anos
16	8	SEXO	C(01)	Sexo, conforme a tabela: 0: Ignorado 1: Masculino 2: Feminino
17	9	RACACOR	C(01)	Raça/Cor: 1: Branca 2: Preta 3: Amarela 4: Parda 5: Indígena
18	10	ESTCIVIL	C(01)	Estado civil, conforme a tabela: 1: Solteiro 2: Casado 3: Viúvo 4: Separado judicialmente 9: Ignorado

19	11	ESC	C(01)	Escolaridade, Anos de estudo concluídos: 1: Nenhuma 2: 1 a 3 anos 3: 4 a 7 anos 4: 8 a 11 anos 5: 12 e mais 9: Ignorado
20	12	OCUP	C(06)	Ocupação, conforme a Classificação Brasileira de Ocupações (CBO-2002)
24	13	CODMUNRES	C(07)	Município de residência do falecido
23	14	CODBAIRES	C(08)	Código do Bairro de residência
26	15	LOCOCOR	C(01)	Local de ocorrência do óbito, conforme a tabela: 9: Ignorado 1: Hospital 2: Outro estab saúde 3: Domicílio 4: Via Pública 5: Outros
27	16	CODESTAB	C(07)	Código do estabelecimento.
31	17	CODMUNOCOR	C(07)	Município de ocorrência do óbito, conforme códigos IBGE.
30	18	CODBAIOCOR	C(08)	Código do bairro de ocorrência.
33	19	IDADEMAE	C(02)	Idade da mãe em anos.
34	20	ESMAE	C(01)	Escolaridade, Anos de estudo concluídos: 1: Nenhuma 2: 1 a 3 anos 3: 4 a 7 anos 4: 8 a 11 anos 5: 12 e mais 9: Ignorado
35	21	OCUPMAE	C(05)	Ocupação da mãe, conforme codificação de OCUPACAO
36	22	QTDFILVIVO	C(02)	Número de filhos vivos.
36	23	QTDFILMORT	C(02)	Número de filhos mortos, ignorados, não incluindo o próprio.
38	24	GRAVIDEZ	C(01)	Tipo de gravidez, conforme a tabela: 9: Ignorado 1: Única 2: Dupla 3: Tripla e mais
37	25	GESTACAO	C(01)	Semanas de gestação, conforme as tabelas: 9: Ignorado 1: Menos de 22 semanas 2: 22 a 27 semanas 3: 28 a 31 semanas 4: 32 a 36 semanas 5: 37 a 41 semanas 6: 42 semanas e mais
39	26	PARTO	C(01)	Tipo de parto, conforme a tabela: 9: Ignorado 1: Vaginal 2: Cesáreo

40	27	OBITOPARTO	C(01)	Morte em relação ao parto, conforme tabela: 9: Ignorado 1: Antes 2: Durante 3: Depois
41	28	PESO	C(04)	Peso ao nascer, em gramas.
42	29	NUMERODN	C(08)	Número da DN
43	30	OBITOGRAV	C(01)	Morte durante a Gravidez conforme tabela: 9: Ignorado 1: Sim 2: Não
44	31	OBITOPUERP	C(01)	Morte durante o puerpério, conforme tabela: 9: Ignorado 1: Sim, ate 42 dias 2: Sim, de 43 dias a 01 ano 3: Não
45	32	ASSISTMED	C(01)	Indica se houve assistência medica, conforme a tabela: 9: Ignorado 1: Com assistência 2: Sem assistência
46	33	EXAME	C(01)	Indica se houve exame complementar, conforme a tabela: 9: Ignorado 1: Sim 2: Não
47	34	CIRURGIA	C(01)	Indica se houve cirurgia, conforme a tabela: 9: Ignorado 1: Sim 2: Não
48	35	NECROPSIA	C(01)	Indica se houve necrópsia, conforme a tabela: 9: Ignorado 1: Sim 2: Não
49	36	LINHAA	C(20)	Linha A do atestado, conforme a Classificação Internacional de Doença (CID), 10a. Revisão.
49	37	LINHAB	C(20)	Linha B do atestado, conforme a Classificação Internacional de Doença (CID), 10a. Revisão.
49	38	LINHAC	C(20)	Linha C do atestado, conforme a Classificação Internacional de Doença (CID), 10a. Revisão.
49	39	LINHAD	C(20)	Linha D do atestado, conforme a Classificação Internacional de Doença (CID), 10a. Revisão.
49	40	LINHA II	C(20)	Linha II do atestado, conforme a Classificação Internacional de Doença (CID), 10a. Revisão.
-	41	CAUSABAS	C(04)	Causa básica, conforme a Classificação Internacional de Doença (CID), 10a. Revisão
	42	TPASSINA	C(01)	-
	43	DTATESTADO	C(08)	Data do Atestado .

56	44	CIRCOBITO	C(01)	Indica o tipo de acidente, se cabível: 9: Ignorado 1: Acidente 2: Suicídio 3: Homicídio 4: Outros
57	45	ACIDTRAB	C(01)	Indica se foi acidente de trabalho, conforme a tabela: 9: Ignorado 1: Sim 2: Não
58	46	FONTE	C(01)	Fonte da informação, conforme a tabela: 9: Ignorado 1: Boletim de Ocorrência 2: Hospital 3: Família 4: Outra
-	47	TPPOS	C(01)	Óbito investigado 1-Sim, 2- Não
-	48	DTINVESTIG	C(08)	Data de investigação.
-	49	CAUSABAS_O	C(04)	Causa básica original, a primeira informação que entra no sistema.
-	50	DTCADASTRO	C(08)	Data de cadastro do registro no sistema.
52	51	ATESTANTE	C(01)	Indica se o medico que assina atendeu o paciente 1: Sim 2: Substituto 3: IML 4: SVO 5: Outros
-	52	FONTEINV	C(01)	Fonte de investigação 1 Comitê de Morte Materna e/ou Infantil 2 Visita domiciliar / Entrevista família 3 Estab Saúde / Prontuário 4 Relacion com outros bancos de dados 5 S V O 6 I M L 7 Outra fonte 8 Múltiplas fontes 9 Ignorado
-	53	DTRECEBIM	C(08)	Data de recebimento no nível central, data da última atualização do registro.
-	54	UFINFORM	C(02)	Código da UF que informou o registro.
-	55	CODINST	C(14)	Código da instalação da geração dos registros.
-	56	CB_PRE	C(04)	Causa selecionada sem re-seleção (novo SCB).

Os campos da seqüência 19 a 29, são preenchidos para óbitos fetais ou para óbitos menores de 1 ano.

**ANEXO D – Descrição da estrutura interna da tabela para o SINASC
(DATASUS, 2008)**

CAMPO FORMULÁRIO	SEQ	NOME	TIPO/TAM	DESCRIÇÃO
Declaração	1	NUMERODN	C(08)	Número da DN, seqüencial por UF informante e por ano.
6	2	LOCNASC	C(01)	Local de ocorrência do nascimento, conforme a tabela: 9: Ignorado 1: Hospital 2: Outro Estab Saúde 3: Domicílio 4: Outros
7	3	CODESTAB	C(07)	Código de estabelecimento de saúde.
10	4	CODBAINASC	C(08)	Código Bairro nascimento.
11	5	CODMUNNASC	C(07)	Código do município de ocorrência.
15	6	IDADEMAE	C(02)	Idade da mãe em anos.
16	7	ESTCIVMAE	C(01)	Estado civil, conforme a tabela: 1: Solteira 2: Casada 3: Viúva 4: Separado judicialmente/Divorciado 9: Ignorado
17	8	ESMAE	C(01)	Escolaridade, anos de estudo concluídos: 1: Nenhuma 2: 1 a 3 anos 3: 4 a 7 anos 4: 8 a 11 anos 5: 12 e mais 9: Ignorado
18	9	CODOCUPMAE	C(06)	Ocupação, conforme a Classificação Brasileira de Ocupações (CBO-2002).
19	10	QTDFILVIVO	C(02)	Número de filhos vivos.
19	11	QTDFILMORT	C(02)	Número de filhos mortos.
22	12	CODBAIRES	C(08)	Código bairro residência.
23	13	CODMUNRES	C(07)	Município de residência da mãe.
25	14	GESTACAO	C(01)	Semanas de gestação, conforme a tabela: 9: Ignorado 1: Menos de 22 semanas 2: 22 a 27 semanas 3: 28 a 31 semanas 4: 32 a 36 semanas 5: 37 a 41 semanas 6: 42 semanas e mais
26	15	GRAVIDEZ	C(01)	Tipo de gravidez, conforme a tabela: 9: Ignorado 1: Única 2: Dupla 3: Tripla e mais

27	16	PARTO	C(01)	Tipo de parto, conforme a tabela: 9: Ignorado 1: Vaginal 2: Cesáreo
28	17	CONSULTAS	C(01)	Número de consultas de pré-natal: 1: Nenhuma 2: de 1 a 3 3: de 4 a 6 4: 7 e mais 9: Ignorado
29	18	DTNASC	C(08)	Data do nascimento, no formato ddmmaaaa
29	19	HORANASC	C(04)	Hora do nascimento
30	20	SEXO	C(01)	Sexo, conforme a tabela: 0: Ignorado 1: Masculino 2: Feminino
31	21	APGAR 1	C(02)	Apgar no primeiro minuto 00 a 10
31	22	APGAR 5	C(02)	Apgar no quinto minuto 00 a 10
32	23	RACACOR	C(01)	Raça/Cor: 1: Branca 2: Preta 3: Amarela 4: Parda 5: Indígena
33	24	PESO	C(04)	Peso ao nascer, em gramas.
34	25	IDANOMAL	C(01)	Anomalia congênita: 9- Ignorado 1= Sim 2= Não
34	26	CODANOMAL	(C20)	Código de malformação congênita ou anomalia cromossômica, de acordo com a CID-10.
-	27	DTCADASTRO	C(08)	Data de cadastramento no sistema.
-	28	DTRECEBIM	C(08)	Data de recebimento no nível central, data da última atualização do registro.
	29	CODINST	C(18)	Código da Instalação da geração dos Registros.
	30	UFINFORM	C(02)	Código da UF que informou o registro.

ANEXO E – Arquivo XML gerado pela Ferramenta *Schema Workbench* do Cenário I.

Na figura 26 é apresentado o Arquivo XML gerado pela Ferramenta *Schema Workbench* através do esquema criado do Cenário I.

```

Schema
<Schema name="SchemaGrupoCausabas">
  <Dimension type="StandardDimension" name="Tempo">
    <Hierarchy hasAll="true" primaryKey="id_tempo">
      <Table name="tbtdimensaotempo" schema="public" alias="tbtdimensaotempo">
      </Table>
      <Level name="Ano" column="ano" type="Integer" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" name="Local">
    <Hierarchy hasAll="true" primaryKey="id_local">
      <Table name="tbtdimensaolocal" schema="public" alias="tbtdimensaolocal">
      </Table>
      <Level name="Estado" column="sigla_estado" type="String" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
      </Level>
      <Level name="Municipio" column="descricao_municipio" type="String" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" name="Causabasica">
    <Hierarchy hasAll="true" primaryKey="id_causabasica">
      <Table name="tbdimensaocausabasica" schema="public">
      </Table>
      <Level name="Grupo" column="descricao_grupo" type="String" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
      </Level>
      <Level name="Causa" column="descr" type="String" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Cube name="Cubo Mortalidade" cache="true" enabled="true">
    <Table name="ctbfatomortalidadecausabasica" schema="public" alias="tbfatomortalidadecausabasica">
    </Table>
    <DimensionUsage source="Tempo" name="Tempo" foreignKey="id_tempo">
    </DimensionUsage>
    <DimensionUsage source="Local" name="Local" foreignKey="id_local">
    </DimensionUsage>
    <DimensionUsage source="Causabasica" name="Causabasica" foreignKey="id_causabasica">
    </DimensionUsage>
    <Measure name="Qtde_Obitos" column="qtde_obitos" formatString="Standard" aggregator="sum" visible="true">
    </Measure>
  </Cube>
</Schema>

```

Figura 26: Arquivo XML Cenário I

ANEXO F – Arquivo *XML* gerado pela Ferramenta *Schema Workbench* do Cenário II.

Na figura 27 é apresentado o Arquivo *XML* gerado pela Ferramenta *Schema Workbench* através do esquema criado do Cenário II.

```

Schema
<Schema name="Indicador_C7">
  <Dimension type="StandardDimension" name="Tempo">
    <Hierarchy hasAll="true" primaryKey="id_tempo">
      <Table name="tbdimensaotempo" schema="public" alias="tbdimensaotempo">
      </Table>
      <Level name="Ano" column="ano" type="Integer" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" name="Local">
    <Hierarchy hasAll="true" primaryKey="id_local">
      <Table name="tbdimensaolocal" schema="public" alias="tbdimensaolocal">
      </Table>
      <Level name="Estado" column="sigla_estado" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
      </Level>
      <Level name="Município" column="descricao_municipio" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Cube name="Sistema_Mortalidade_DoencasInfeccaoRespiratoriaAguda" cache="true" enabled="true">
    <Table name="cfatomortalidademenores5anosinfeccaorespc7" schema="public" alias="tbfatomortalidademenores5anosinfeccaorespiratoriaAguda_IndicadorC7">
    </Table>
    <DimensionUsage source="Tempo" name="Tempo" foreignKey="id_tempo">
    </DimensionUsage>
    <DimensionUsage source="Local" name="Local" foreignKey="id_local">
    </DimensionUsage>
    <Measure name="Nro_Causas" column="qtde_obitos" aggregator="sum" visible="true">
    </Measure>
  </Cube>
  <Cube name="Sistema_Mortalidade_DoencasInfeccaoRespiratoriaAguda" cache="true" enabled="true">
    <Table name="cfatomortalidademenores5anosinfeccaorespc7" schema="public" alias="tbfatomortalidademenores5anosinfeccaorespiratoriaAguda_IndicadorC7">
    </Table>
    <DimensionUsage source="Tempo" name="Tempo" foreignKey="id_tempo">
    </DimensionUsage>
    <DimensionUsage source="Local" name="Local" foreignKey="id_local">
    </DimensionUsage>
    <Measure name="Nro_Causas" column="qtde_obitos" aggregator="sum" visible="true">
    </Measure>
  </Cube>
  <Cube name="Sistema_MortalidadeMenores5Anos" cache="true" enabled="true">
    <Table name="cfatomortalidademenores5anos16" schema="public" alias="tbfatomortalidademenores5anos">
    </Table>
    <DimensionUsage source="Tempo" name="Tempo" foreignKey="id_tempo">
    </DimensionUsage>
    <DimensionUsage source="Local" name="Local" foreignKey="id_local">
    </DimensionUsage>
    <Measure name="Nro_Obitos" column="qtde_obitos" aggregator="sum" visible="true">
    </Measure>
  </Cube>
  <VirtualCube enabled="true" name="C7_Mortalidade_Proporcional_Por_InfeccaoRespAguda_Menores5Anos">
    <VirtualCubeDimension name="Tempo">
    </VirtualCubeDimension>
    <VirtualCubeDimension name="Local">
    </VirtualCubeDimension>
    <VirtualCubeMeasure cubeName="Sistema_Mortalidade_DoencasInfeccaoRespiratoriaAguda" name="[Measures].[Nro_Causas]" visible="true">
    </VirtualCubeMeasure>
    <VirtualCubeMeasure cubeName="Sistema_MortalidadeMenores5Anos" name="[Measures].[Nro_Obitos]" visible="true">
    </VirtualCubeMeasure>
    <CalculatedMember name="IndicadorC7_MortalidadeProporcional_InfeccaoRespAguda_Menores5Anos" formula="([Measures].[Nro_Causas] / [Measures].[Nro_Obitos]) * 100"
    dimension="Measures" visible="true">
    </CalculatedMember>
  </VirtualCube>
</Schema>

```

Figura 27: Barra Arquivo XML Cenário II

ANEXO G – Causas Básicas por Doenças do Aparelho Circulatório na Tabela CID-10

A tabela 10 apresenta as doenças do Aparelho Circulatório cadastradas na CID-10, utilizadas no Cenário I. Na primeira e terceira colunas são apresentadas os códigos e ao lado respectivamente sua descrição.

Tabela 10: Tabela das Causas Básicas por Doenças do Aparelho Circulatório

cid10	Descrição	Cid10	Descrição
I00	I00 Febre reumatica s/mencao de compr do coracao	I499	I49.9 Arritmia cardiaca NE
I01	I01 Febre reumatica c/compr do coracao	I50	I50 Insuf cardiaca
I010	I01.0 Pericardite reumatica aguda	I500	I50.0 Insuf cardiaca congestiva
I011	I01.1 Endocardite reumatica aguda	I501	I50.1 Insuf ventricular esquerda
I012	I01.2 Miocardite reumatica aguda	I509	I50.9 Insuf cardiaca NE
I018	I01.8 Outr form de doenc cardiaca reumatica aguda	I51	I51 Complic cardiopatas doenc cardiacas mal def
I019	I01.9 Doenc cardiaca reumatica aguda NE	I510	I51.0 Defeito adquir de septo cardiaco
I02	I02 Coreia reumatica	I511	I51.1 Ruptura de cordoalha tendinea NCOP
I020	I02.0 Coreia reumatica c/compr cardiaco	I512	I51.2 Ruptura de musculo papilar NCOP
I029	I02.9 Coreia reumatica s/compr cardiaco	I513	I51.3 Trombose intracardiaca NCOP
I05	I05 Doenc reumaticas da valva mitral	I514	I51.4 Miocardite NE
I050	I05.0 Estenose mitral	I515	I51.5 Degeneracao miocardica
I051	I05.1 Insuf mitral reumatica	I516	I51.6 Doenc cardiovascular NE
I052	I05.2 Estenose mitral c/insuf	I517	I51.7 Cardiomegalia
I058	I05.8 Outr doenc da valva mitral	I518	I51.8 Outr doenc mal definidas do coracao
I059	I05.9 Doenc NE da valva mitral	I519	I51.9 Doenc NE do coracao
I06	I06 Doenc reumaticas da valva aortica	I52	I52 Outr afecoes cardiacas em doenc COP
I060	I06.0 Estenose aortica reumatica	I520	I52.0 Outr afecoes cardiacas em doenc bacter COP
I061	I06.1 Insuf aortica reumatica	I521	I52.1 Outr afecc card outr doenc infec parasit COP
I062	I06.2 Estenose aortica reumatica c/insuf	I528	I52.8 Outr afecoes cardiacas em outr doenc COP
I068	I06.8 Outr doenc reumaticas da valva aortica	I60	I60 Hemorragia subaracnoide
I069	I06.9 Doenc reumatica da valva aortica NE	I600	I60.0 Hemorragia subarac prov sifao bifurc carotid
I07	I07 Doenc reumaticas da valva tricuspide	I601	I60.1 Hemorragia subarac prov arteria cerebr media
I070	I07.0 Estenose tricuspide	I602	I60.2 Hemorragia subarac prov arteria comunic ant
I071	I07.1 Insuf tricuspide	I603	I60.3 Hemorragia subarac prov arteria comunic post
I072	I07.2 Estenose tricuspide c/insuf	I604	I60.4 Hemorragia subarac prov arteria basilar
I078	I07.8 Outr doenc da valva tricuspide	I605	I60.5 Hemorragia subarac prov arteria vertebral
I079	I07.9 Doenc NE da valva tricuspide	I606	I60.6 Hemorragia subarac prov outr arter intracran
I08	I08 Doenc de mult valvas	I607	I60.7 Hemorragia subarac prov arteria intracran NE
I080	I08.0 Transt de valvas mitral e aortica	I608	I60.8 Outr hemorragias subaracnoides
I081	I08.1 Transt de ambas as valvas mitral tricuspide	I609	I60.9 Hemorragia subaracnoide NE

I082	I08.2 Transt de ambas as valvas aortica tricuspide	I61	I61 Hemorragia intracerebral
I083	I08.3 Transt das valvas mitral aortica tricuspide	I610	I61.0 Hemorragia intracerebr hemisf subcortical
I088	I08.8 Outr doenc de mult valvas	I611	I61.1 Hemorragia intracerebr hemisferica cortical
I089	I08.9 Doenc NE de mult valvas	I612	I61.2 Hemorragia intracerebral hemisferica NE
I09	I09 Outr doenc reumaticas do coracao	I613	I61.3 Hemorragia intracerebral do tronco cerebral
I090	I09.0 Miocardite reumatica	I614	I61.4 Hemorragia intracerebral cerebelar
I091	I09.1 Doenc reumaticas do endocardio valva NE	I615	I61.5 Hemorragia intracerebral intraventricular
I092	I09.2 Pericardite reumatica cronica	I616	I61.6 Hemorragia intracerebral de mult localiz
I098	I09.8 Outr doenc reumaticas espec do coracao	I618	I61.8 Outr hemorragias intracerebrais
I099	I09.9 Doenc cardiaca reumatica NE	I619	I61.9 Hemorragia intracerebral NE
I10	I10 Hipertensao essencial	I62	I62 Outr hemorragias intracranianas nao-traum
I11	I11 Doenc cardiaca hipertensiva	I620	I62.0 Hemorragia subdural
I110	I11.0 Doenc cardiaca hipertensiva c/insuf cardiaca	I621	I62.1 Hemorragia extradural nao-traum
I119	I11.9 Doenc cardiaca hipertensiva s/insuf cardiaca	I629	I62.9 Hemorragia intracraniana NE
I12	I12 Doenc renal hipertensiva	I63	I63 Infarto cerebral
I120	I12.0 Doenc renal hipertensiva c/insuf renal	I630	I63.0 Infarto cerebr dev tromb arterias pre-cerebr
I129	I12.9 Doenc renal hipertensiva s/insuf renal	I631	I63.1 Infarto cerebr dev embol arterias pre-cerebr
I13	I13 Doenc cardiaca e renal hipertensiva	I632	I63.2 Inf cer dev oclusao esten NE art pre-cerebr
I130	I13.0 Doenc cardiaca e renal hipert c/insufic card	I633	I63.3 Infarto cerebr dev trombose arterias cerebr
I131	I13.1 Doenc cardiaca e renal hipert c/insuf renal	I634	I63.4 Infarto cerebral dev embolia arterias cerebr
I132	I13.2 Doenc card renal hipert c/insuf card e renal	I635	I63.5 Inf cer dev oclusao estenose NE art cerebr
I139	I13.9 Doenc cardiaca e renal hipertensiva NE	I636	I63.6 Inf cerebr dev tromb ven cerebr nao-piogenic
I15	I15 Hipertensao secund	I638	I63.8 Outr infartos cerebrais
I150	I15.0 Hipertensao renovascular	I639	I63.9 Infarto cerebral NE
I151	I15.1 Hipertensao secund a outr afecoes renais	I64	I64 Acid vasc cerebr NE como hemorrag isquemico
I152	I15.2 Hipertensao secund a afecoes endocrinas	I65	I65 Oclus/esten art pre-cerebr q n res inf cerebr
I158	I15.8 Outr form de hipertensao secund	I650	I65.0 Oclusao e estenose da arteria vertebral
I159	I15.9 Hipertensao secund NE	I651	I65.1 Oclusao e estenose da arteria basilar
I20	I20 Angina pectoris	I652	I65.2 Oclusao e estenose da arteria carotida
I200	I20.0 Angina instavel	I653	I65.3 Oclusao/estenose art pre-cerebr mult bilat
I201	I20.1 Angina pectoris c/espasmo documentado	I658	I65.8 Oclusao e estenose outr arteria pre-cerebral
I208	I20.8 Outr form de angina pectoris	I659	I65.9 Oclusao e estenose arterias pre-cerebrais NE
I209	I20.9 Angina pectoris NE	I66	I66 Oclusao/estenose art cerebr q n res inf cerebr
I21	I21 Infarto agudo do miocardio	I660	I66.0 Oclusao e estenose da arteria cerebral media
I210	I21.0 Infarto agudo transmural parede ant miocard	I661	I66.1 Oclusao e estenose arteria cerebral anterior
I211	I21.1 Infarto agudo transmural parede inf miocard	I662	I66.2 Oclusao e estenose arteria cerebral post
I212	I21.2 Infarto agudo transmural miocardio outr loc	I663	I66.3 Oclusao e estenose de arterias cerebelares
I213	I21.3 Infarto agudo transmural miocardio loc NE	I664	I66.4 Oclusao/estenose arterias cerebr mult bilat
I214	I21.4 Infarto agudo subendocardico do miocardio	I668	I66.8 Oclusao e estenose de outr arteria cerebral
I219	I21.9 Infarto agudo do miocardio NE	I669	I66.9 Oclusao e estenose de arteria cerebral NE
I22	I22 Infarto do miocardio recorrente	I67	I67 Outr doenc cerebrovasculares
I220	I22.0 Infarto do miocardio recorrente parede ant	I670	I67.0 Dissecacao de arterias cerebrais s/ruptura
I221	I22.1 Infarto do miocardio recorrente parede infer	I671	I67.1 Aneurisma cerebral nao-roto

I228	I22.8 Infarto do miocardio recorrente outr localiz	I672	I67.2 Aterosclerose cerebral
I229	I22.9 Infarto do miocardio recorrente localiz NE	I673	I67.3 Leucoencefalopatia vascular progressiva
I23	I23 Alg complic atuais subs infarto agud miocard	I674	I67.4 Encefalopatia hipertensiva
I230	I23.0 Hemoperic c/compl subs infart agud miocard	I675	I67.5 Doenc de Moyamoya
I231	I23.1 Com interatr c/comp atual subs inf agud mioc	I676	I67.6 Trombose nao-piogenica sist venoso intracran
I232	I23.2 Com interven c/comp atual subs inf agud mioc	I677	I67.7 Arterite cerebral NCOP
I233	I23.3 Rupt par cor s/hemop c/comp subs inf ag mioc	I678	I67.8 Outr doenc cerebrovasculares espec
I234	I23.4 Rupt cord tend c/comp atual subs inf ag mioc	I679	I67.9 Doenc cerebrovascular NE
I235	I23.5 Rupt musc pap c/comp atual subs inf ag mioc	I68	I68 Transt cerebrovasculares em doenc COP
I236	I23.6 Tromb atr aur/vent c/comp at sub inf ag mioc	I680	I68.0 Angiopatia cerebral amiloidotica
I238	I23.8 Outr complic atuais subs infart agud miocard	I681	I68.1 Arterite cerebral em doenc infecc e parasit
I24	I24 Outr doenc isquemias agudas do coracao	I682	I68.2 Arterite cerebral em outr doenc COP
I240	I24.0 Tromb coronaria q n result infarto miocardio	I688	I68.8 Outr transt cerebrovasculares em doenc COP
I241	I24.1 Sindr de Dressler	I69	I69 Sequelas de doenc cerebrovasculares
I248	I24.8 Outr form doenc isquemica aguda do coracao	I690	I69.0 Sequelas de hemorragia subaranoidea
I249	I24.9 Doenc isquemica aguda do coracao NE	I691	I69.1 Sequelas de hemorragia intracerebral
I25	I25 Doenc isquemica cronica do coracao	I692	I69.2 Sequelas outr hemorrag intracran nao traum
I250	I25.0 Doenc cardiovasc ateroscler descr deste modo	I693	I69.3 Sequelas de infarto cerebral
I251	I25.1 Doenc aterosclerotica do coracao	I694	I69.4 Sequelas acid vasc cerebr NE c/hemorr isquem
I252	I25.2 Infarto antigo do miocardio	I698	I69.8 Sequelas outr doenc cerebrovasculares e NE
I253	I25.3 Aneurisma cardiaco	I70	I70 Aterosclerose
I254	I25.4 Aneurisma de arteria coronaria	I700	I70.0 Aterosclerose da aorta
I255	I25.5 Miocardiopatia isquemica	I701	I70.1 Aterosclerose da arteria renal
I256	I25.6 Isquemia miocardica silenciosa	I702	I70.2 Aterosclerose das arterias das extremidades
I258	I25.8 Outr form de doenc isquemica cronica coracao	I708	I70.8 Aterosclerose de outr arterias
I259	I25.9 Doenc isquemica cronica do coracao NE	I709	I70.9 Aterosclerose generalizada e a NE
I26	I26 Embolia pulmonar	I71	I71 Aneurisma e disseccao da aorta
I260	I26.0 Embolia pulmonar c/menc cor pulmonale agudo	I710	I71.0 Aneurisma dissecante da aorta
I269	I26.9 Embolia pulmonar s/menc cor pulmonale agudo	I711	I71.1 Aneurisma da aorta toracica roto
I27	I27 Outr form de doenc cardiaca pulmonar	I712	I71.2 Aneurisma da aorta toracica s/mencao ruptura
I270	I27.0 Hipertensao pulmonar prim	I713	I71.3 Aneurisma da aorta abdominal roto
I271	I27.1 Cardiopatia cifoescoliotica	I714	I71.4 Aneurisma aorta abdominal s/mencao ruptura
I278	I27.8 Outr doenc pulmonares do coracao espec	I715	I71.5 Aneurisma da aorta toraco-abdominal roto
I279	I27.9 Cardiopatia pulmonar NE	I716	I71.6 Aneurisma aorta toraco-abdom s/menc ruptura
I28	I28 Outr doenc dos vasos pulmonares	I718	I71.8 Aneurisma da aorta de localiz NE roto
I280	I28.0 Fistula arteriovenosa de vasos pulmonares	I719	I71.9 Aneurisma aortico local NE s/mencao ruptura
I281	I28.1 Aneurisma da arteria pulmonar	I72	I72 Outr aneurismas
I288	I28.8 Outr doenc espec dos vasos pulmonares	I720	I72.0 Aneurisma da arteria carotida
I289	I28.9 Doenc NE dos vasos pulmonares	I721	I72.1 Aneurisma de arteria dos membros super
I30	I30 Pericardite aguda	I722	I72.2 Aneurisma da arteria renal
I300	I30.0 Pericardite aguda idiopatica nao especifica	I723	I72.3 Aneurisma de arteria iliaca
I301	I30.1 Pericardite infecc	I724	I72.4 Aneurisma de arteria dos membros infer
I308	I30.8 Outr form de pericardite aguda	I728	I72.8 Aneurisma de outr arterias espec

I309	I30.9 Pericardite aguda NE	I729	I72.9 Aneurisma de localiz NE
I31	I31 Outr doenc do pericardio	I73	I73 Outr doenc vasculares perifericas
I310	I31.0 Pericardite adesiva cronica	I730	I73.0 Sindr de Raynaud
I311	I31.1 Pericardite constrictiva cronica	I731	I73.1 Tromboangeite obliterante
I312	I31.2 Hemopericardio NCOP	I738	I73.8 Outr doenc vasculares perifericas espec
I313	I31.3 Derrame pericardico	I739	I73.9 Doenc vasculares perifericas NE
I318	I31.8 Outr doenc espec do pericardio	I74	I74 Embolia e trombose arteriais
I319	I31.9 Doenc NE do pericardio	I740	I74.0 Embolia e trombose da aorta abdominal
I32	I32 Pericardite em doenc COP	I741	I74.1 Embolia trombose outr porcoes da aorta e NE
I320	I32.0 Pericardite em doenc bacter COP	I742	I74.2 Embolia e trombose de arterias membros super
I321	I32.1 Pericardite outr doenc infecc e parasit COP	I743	I74.3 Embolia e trombose de arterias membros infer
I328	I32.8 Pericardite em outr doenc COP	I744	I74.4 Embolia e trombose de arterias membros NE
I33	I33 Endocardite aguda e subaguda	I745	I74.5 Embolia e trombose da arteria iliaca
I330	I33.0 Endocardite infecc aguda e subaguda	I748	I74.8 Embolia e trombose de outr arterias
I339	I33.9 Endocardite aguda NE	I749	I74.9 Embolia e trombose de arteria NE
I34	I34 Transt nao-reumaticos da valva mitral	I77	I77 Outr afeccoes das arterias e arteriolas
I340	I34.0 Insuf mitral	I770	I77.0 Fistula arteriovenosa adquir
I341	I34.1 Prolapso mitral	I771	I77.1 Estenose de arteria
I342	I34.2 Estenose mitral nao-reumatica	I772	I77.2 Ruptura de arteria
I348	I34.8 Outr transt nao-reumaticos da valva mitral	I773	I77.3 Displasia fibromuscular arterial
I349	I34.9 Transt nao-reumaticos da valva mitral NE	I774	I77.4 Sindr de compressao da arteria celiaca
I35	I35 Transt nao-reumaticos da valva aortica	I775	I77.5 Necrose de arteria
I350	I35.0 Estenose aortica	I776	I77.6 Arterite NE
I351	I35.1 Insuf aortica	I778	I77.8 Outr afeccoes espec arterias e arteriolas
I352	I35.2 Estenose aortica c/insuf	I779	I77.9 Afeccoes de arterias e arteriolas NE
I358	I35.8 Outr transt da valva aortica	I78	I78 Doenc dos capilares
I359	I35.9 Transt NE da valva aortica	I780	I78.0 Telangiectasia hemorragica hereditaria
I36	I36 Transt nao-reumaticos da valva tricuspide	I781	I78.1 Nevo nao-neoplasico
I360	I36.0 Estenose tricuspide nao-reumatica	I788	I78.8 Outr doenc dos capilares
I361	I36.1 Insuf tricuspide nao-reumatica	I789	I78.9 Doenc NE dos capilares
I362	I36.2 Estenose tricuspide c/insuf nao-reumatica	I79	I79 Transt arter arteriolas capilares doenc COP
I368	I36.8 Outr transt nao-reumaticos valva tricuspide	I790	I79.0 Aneurisma da aorta em doenc COP
I369	I36.9 Transt nao-reumaticos da valva tricuspide NE	I791	I79.1 Aortite em doenc COP
I37	I37 Transt da valva pulmonar	I792	I79.2 Angiopatia periferica em doenc COP
I370	I37.0 Estenose da valva pulmonar	I798	I79.8 Outr transt arter arteriolas capil doenc COP
I371	I37.1 Insuf da valva pulmonar	I80	I80 Flebite e tromboflebite
I372	I37.2 Estenose da valva pulmonar c/insuf	I800	I80.0 Flebite tromboflebite vasos superf membr inf
I378	I37.8 Outr transt da valva pulmonar	I801	I80.1 Flebite e tromboflebite da veia femural
I379	I37.9 Transt NE da valva pulmonar	I802	I80.2 Flebite tromboflebit outr vasos prof membr inf
I38	I38 Endocardite de valva NE	I803	I80.3 Flebite e tromboflebite dos membros infer NE
I39	I39 Endocardite transt valvulares card doenc COP	I808	I80.8 Flebite e tromboflebite de outr localiz
I390	I39.0 Transt da valva mitral em doenc COP	I809	I80.9 Flebite e tromboflebite de localiz NE
I391	I39.1 Transt da valva aortica em doenc COP	I81	I81 Trombose da veia porta

I392	I39.2 Transt da valva tricuspide em doenc COP	I82	I82 Outr embolia e trombose venosas
I393	I39.3 Transt da valva pulmonar em doenc COP	I820	I82.0 Sindr de Budd-Chiari
I394	I39.4 Transt de mult valvas em doenc COP	I821	I82.1 Tromboflebite migratoria
I398	I39.8 Endocardite de valva NE em doenc COP	I822	I82.2 Embolia e trombose de veia cava
I40	I40 Miocardite aguda	I823	I82.3 Embolia e trombose de veia renal
I400	I40.0 Miocardite infecc	I828	I82.8 Embolia e trombose de outr veias espec
I401	I40.1 Miocardite isolada	I829	I82.9 Embolia e trombose venosas de veia NE
I408	I40.8 Outr miocardites agudas	I83	I83 Varizes dos membros infer
I409	I40.9 Miocardite aguda NE	I830	I83.0 Varizes dos membros infer c/ulcera
I41	I41 Miocardite em doenc COP	I831	I83.1 Varizes dos membros infer c/inflam
I410	I41.0 Miocardite em doenc bacter COP	I832	I83.2 Varizes membros infer c/ulcera e inflam
I411	I41.1 Miocardite em doenc virais COP	I839	I83.9 Varizes membros infer s/ulcera ou inflam
I412	I41.2 Miocardite outr doenc infecc e parasit COP	I84	I84 Hemorroidas
I418	I41.8 Miocardite em outr doenc COP	I840	I84.0 Hemorroidas internas trombosadas
I42	I42 Cardiomiopatis	I841	I84.1 Hemorroidas internas c/outr complic
I420	I42.0 Cardiomiopatia dilatada	I842	I84.2 Hemorroidas internas s/complic
I421	I42.1 Cardiomiopatia obstrutiva hipertrofica	I843	I84.3 Hemorroidas externas trombosadas
I422	I42.2 Outr cardiomiopatis hipertroficas	I844	I84.4 Hemorroidas externas c/outr complic
I423	I42.3 Doenc endomiocardica	I845	I84.5 Hemorroidas externas s/complic
I424	I42.4 Fibroelastose endocardica	I846	I84.6 Plicomas hemorroidarios residuais
I425	I42.5 Outr cardiomiopatis restritivas	I847	I84.7 Hemorroidas trombosadas NE
I426	I42.6 Cardiomiopatia alcoolica	I848	I84.8 Hemorroidas NE c/outr complic
I427	I42.7 Cardiomiopatia dev drogas e outr agentes ext	I849	I84.9 Hemorroidas s/complic NE
I428	I42.8 Outr cardiomiopatis	I85	I85 Varizes esofagianas
I429	I42.9 Cardiomiopatia NE	I850	I85.0 Varizes esofagianas sangrantes
I43	I43 Cardiomiopatia em doenc COP	I859	I85.9 Varizes esofagianas s/sangramento
I430	I43.0 Cardiomiopatia em doenc infecc e parasit COP	I86	I86 Varizes de outr localiz
I431	I43.1 Cardiomiopatia em doenc metabolicas	I860	I86.0 Varizes sublinguais
I432	I43.2 Cardiomiopatia em doenc nutricionais	I861	I86.1 Varizes escrotais
I438	I43.8 Cardiomiopatia em outr doenc COP	I862	I86.2 Varizes pelvicas
I44	I44 Bloqueio atrioventricular e do ramo esquerdo	I863	I86.3 Varizes da vulva
I440	I44.0 Bloqueio atrioventricular de 1.grau	I864	I86.4 Varizes gastricas
I441	I44.1 Bloqueio atrioventricular de 2.grau	I868	I86.8 Varizes de outr localiz espec
I442	I44.2 Bloqueio atrioventricular total	I87	I87 Outr transt das veias
I443	I44.3 Outr form de bloqueio atrioventricular e NE	I870	I87.0 Sindr pos-flebite
I444	I44.4 Bloqueio do fasciculo anterior esquerdo	I871	I87.1 Compressao venosa
I445	I44.5 Bloqueio do fasciculo posterior esquerdo	I872	I87.2 Insuf venosa
I446	I44.6 Outr form de bloqueio fascicular e as NE	I878	I87.8 Outr transt venosos espec
I447	I44.7 Bloqueio de ramo esquerdo NE	I879	I87.9 Transt venoso NE
I45	I45 Outr transt de conducao	I88	I88 Linfadenite inespecifica
I450	I45.0 Bloqueio fascicular direito	I880	I88.0 Linfadenite mesenterica nao especifica
I451	I45.1 Outr form de bloqueio de ramo direito e NE	I881	I88.1 Linfadenite cronica exceto a mesenterica
I452	I45.2 Bloqueio bifascicular	I888	I88.8 Outr linfadenites inespecificas

I453	I45.3 Bloqueio trifascicular	I889	I88.9 Linfadenite NE
I454	I45.4 Bloqueio intraventricular nao especifico	I89	I89 Outr transt nao-infecc vasos linf gangl linf
I455	I45.5 Outr form espec de bloqueio cardiaco	I890	I89.0 Linfedema NCOP
I456	I45.6 Sindr de pre-excitacao	I891	I89.1 Linfangite
I458	I45.8 Outr transt espec da conducao	I898	I89.8 Out transt n-inf espec vasos linf gangl linf
I459	I45.9 Transt de conducao NE	I899	I89.9 Transt nao-infecc vasos linf gangl linf NE
I46	I46 Parada cardiaca	I95	I95 Hipotensao
I460	I46.0 Parada cardiaca c/ressuscitacao bem sucedida	I950	I95.0 Hipotensao idiopatica
I461	I46.1 Morte subita cardiaca descrita desta form	I951	I95.1 Hipotensao ortostatica
I469	I46.9 Parada cardiaca NE	I952	I95.2 Hipotensao dev drogas
I47	I47 Taquicardia paroxistica	I958	I95.8 Outr tipos de hipotensao
I470	I47.0 Arritmia ventricular p/reentrada	I959	I95.9 Hipotensao NE
I471	I47.1 Taquicardia supraventricular	I97	I97 Transt aparelho circulat subseq proced NCOP
I472	I47.2 Taquicardia ventricular	I970	I97.0 Sindr pos-cardiotomia
I479	I47.9 Taquicardia paroxistica NE	I971	I97.1 Outr disturb func subseq cirurgia cardiaca
I48	I48 Flutter e fibrilacao atrial	I972	I97.2 Sindr do linfedema pos-mastectomia
I49	I49 Outr arritmias cardiacas	I978	I97.8 Outr transt ap circulat subseq proced NCOP
I490	I49.0 Flutter e fibrilacao ventricular	I979	I97.9 Transt aparelho circulat subseq proced NE
I491	I49.1 Despolarizacao atrial prematura	I98	I98 Outr transt aparelho circulatorio doenc COP
I492	I49.2 Despolarizacao juncional prematura	I980	I98.0 Sifilis cardiovascular
I493	I49.3 Despolarizacao ventricular prematura	I981	I98.1 Transt cardiovasc outr doenc inf parasit COP
I494	I49.4 Outr form de despolarizacao prematura e NE	I982	I98.2 Varizes esofagianas em doenc COP
I495	I49.5 Sindr do no sinusal	I988	I98.8 Outr transt espec ap circulat doenc COP
I498	I49.8 Outr arritmias cardiacas espec	I99	I99 Outr transt do aparelho circulatorio e os NE