UNIVERSIdade FEEVALE

PAULO ALBERTO COUTINHO PRECHT

*BUSINESS INTELLIGENCE* E *DATA WAREHOUSE* APLICADOs A PESQUISA NACIONAL DE SAÚDE DE 2013

Novo Hamburgo

2016

PAULO ALBERTO COUTINHO PRECHT

*BUSINESS INTELLIGENCE* E *DATA WAREHOUSE* APLICADOs A PESQUISA NACIONAL DE SAÚDE DE 2013

Trabalho de Conclusão de Curso

apresentado como requisito parcial

à obtenção do grau de Bacharel em

Sistemas de Informação pela

Universidade Feevale

Orientador: Dr. Juliano Varella de Carvalho

Novo Hamburgo

2016

Agradecimentos

Agradeço a todos que de alguma maneira contribuíram e me acompanharam ao longo da realização deste trabalho de conclusão e também de todo o curso de Sistemas de Informação.

Um agradecimento em especial ao meu orientador Dr. Juliano Varella de Carvalho, aos meus amigos e principalmente a minha família.

A todos meu muito obrigado.

Resumo

Os governos municipais e estaduais, além do governo federal, divulgam em seus portais públicos e abertos diversos dados e informações da área da saúde, de maneira a facilitar a compreensão para os gestores da área da saúde e para a sociedade em geral. Porém, a forma como os dados são fornecidos é precária e não expõe as informações de forma analítica. Normalmente os dados são apresentados de maneira estática e pouco intuitiva, o que acaba dificultando a leitura e compreensão destas informações. Este estudo teve como objetivo descobrir e identificar fontes de dados relevantes e de interesse social nestes portais e através de ferramentas de *Business Intelligence* e *Data Warehouse*, foram desenvolvidos cubos e visões dentro de um ambiente que possibilitou realizar análises e cruzamentos, além de exibir indicadores e gráficos dos dados obtidos. Desta forma, este trabalho constatou a viabilidade do uso de dados do DATASUS para construir visualizações intuitivas e dinâmicas, que podem ser utilizadas por profissionais da área da saúde.

Palavras-chave: *Business Intelligence*. *Data warehouse*. *OLAP*. Sistema de Saúde Brasileiro.

Abstract

The State and local governments as well as the federal government, disclose in their public and open portals, various data and health information in order to facilitate understanding for managers of health and society in general. However, the way the data is provided is poor and does not expose the information analytically. Normally the data are presented in static and unintuitive way, which makes it difficult to read and understand this information. This study aimed to discover and identify relevant data sources and social interest in these portals and through tools of Business Intelligence and Data Warehouse were developed cubes and views in an environment that made it possible to perform analysis and crossings, and also to display indicators and graphics data. Thus, this study verified the feasibility of the DATASUS data use to build intuitive and dynamic views, which can be used by health professionals.

Key words: Business Intelligence. Data warehouse. OLAP. Brazilian Health System.

Lista de Figuras

[Figura 1 - Estrutura de um *Data Warehouse* e seus *Data Marts* 17](#_Toc455940809)

[Figura 2 - Processo de *ETL* 18](#_Toc455940810)

[Figura 3 - Granularidades diferentes em um mesmo assunto: Vendas 19](#_Toc455940811)

[Figura 4 - Fato compra e seus elementos 20](#_Toc455940812)

[Figura 5 - Estrutura de um Modelo Estrela (*Star schema*) 21](#_Toc455940813)

[Figura 6 - Estrutura de um Modelo Floco de Neve (*Snowflake*) 22](#_Toc455940814)

[Figura 7 - Cubo *OLAP*: Vendas 23](#_Toc455940815)

[Figura 8 - Relatórios interativos no Pentaho 28](#_Toc455940816)

[Figura 9 - Oracle *Business Intelligence* exemplo de uma *dashboard* 29](#_Toc455940817)

[Figura 10 – Tela inicial da interface do usuário final do IBM Cognos 31](#_Toc455940818)

[Figura 11 - Arquivo de microdados “PESPNS2013.txt” 40](#_Toc455940819)

[Figura 12 – Resumo das etapas: Extração e criação do *Data Warehouse*; Transformação e Carga 41](#_Toc455940820)

[Figura 13 - Arquivo microdados.zip descompactado 41](#_Toc455940821)

[Figura 14 – Trecho do Dicionário de variáveis do arquivo de microdados da PNS 2013 42](#_Toc455940822)

[Figura 15 - Comando *Select* na tabela "pes\_dados" 44](#_Toc455940823)

[Figura 16 - Comando *Select* na tabela "pes\_variaveis" 45](#_Toc455940824)

[Figura 17 - Comando *Select* na tabela “pes\_resplegend\_p05404” 45](#_Toc455940825)

[Figura 18 - *Schema* “pns” 46](#_Toc455940826)

[Figura 19 – Modelo estrela da *view* “fato\_acidentes” e suas dimensões 48](#_Toc455940827)

[Figura 20 – Modelo estrela da *view* “fato\_uso\_alcool” e suas dimensões 50](#_Toc455940828)

[Figura 21 - Etapas da preparação e instalação do *Pentaho Community* 52](#_Toc455940829)

[Figura 22 - Diretório do *Pentaho Community* 54](#_Toc455940830)

[Figura 23 - Desenvolvimento: Tela inicial do *Pentaho Community* 55](#_Toc455940831)

[Figura 24 - *Pentaho Community*: conexão com o banco de dados com usuário root 56](#_Toc455940832)

[Figura 25 - Seleção das tabelas fato e dimensões com usuário root 57](#_Toc455940833)

[Figura 26 - Seleção das tabelas fato e dimensões com o usuário limitado 58](#_Toc455940834)

[Figura 27 – Criação dos *joins* do cubo “acidentes” 59](#_Toc455940835)

[Figura 28 – *Pentaho Community*: seleção e configuração das medidas do cubo 60](#_Toc455940836)

[Figura 29 - Restrição de realizar um único *join* entre a tabela fato e uma tabela dimensão no *Pentaho Community* 64](#_Toc455940837)

[Figura 30 - *Plug-ins* adicionais de OLAP para o *Pentaho Community* 65](#_Toc455940838)

[Figura 31 – Todas as dimensões e medidas do cubo “acidentes” 66](#_Toc455940839)

[Figura 32 - Visualização OLAP do cubo “acidentes” (“Qtd. entrevistados” x “Se envolveram em acid. c/lesões ult. 12m” x “Sexo” x “Estado civil”) 67](#_Toc455940840)

[Figura 33 – Gráfico OLAP do cubo “acidentes” (“Qtd. entrevistados” x “Se envolveram em acid. c/lesões ult. 12m” = sim x “Sexo” x “Estado civil”) 68](#_Toc455940841)

[Figura 34 - Visualização e gráfico OLAP do cubo “acidentes” (“Qtd. entrevistados” x “Se envolveram em acid. c/lesões ult. 12m” x “Sexo” x “Estado civil” x “Faixa etária”) 69](#_Toc455940842)

[Figura 35 - Visualização OLAP do cubo “uso\_alcool” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” x “Sexo” x “Estado civil”) 70](#_Toc455940843)

[Figura 36 - Visualização OLAP do cubo “uso\_alcool” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” x “Sexo” x “Estado civil” x “Qtd. dias na semana que costumam consumir beb. alc.”) 71](#_Toc455940844)

[Figura 37 – Gráfico do cubo “uso\_alcool” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” x “Sexo” x “Estado civil” x “Qtd. dias na semana que costumam consumir beb. alc.”) 71](#_Toc455940845)

[Figura 38 - Modelo estrela da *view* “fato\_acid\_vs\_uso\_alc” e suas dimensões 72](#_Toc455940846)

[Figura 39 - Visualização OLAP do cubo “acid\_vs\_uso\_alc” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” x “Se envolveram em acid. c/lesões ult. 12m” x “Estado civil”) 75](#_Toc455940847)

[Figura 40 - Visualização OLAP do cubo “acid\_vs\_uso\_alc” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” = “Sim” x “Se envolveram em acid. c/lesões ult. 12m” = “Sim” x “Estado civil” x “Faixa etária”) 76](#_Toc455940848)

[Figura 41 - Gráfico do cubo “acid\_vs\_uso\_alc” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” = “Sim” x “Se envolveram em acid. c/lesões ult. 12m” = “Sim” x “Estado civil” = “Solteiro” x “Faixa etária”) 77](#_Toc455940849)

[Figura 42 - *Script*: 4-aux\_faixa\_etaria.sql 88](#_Toc455940850)

[Figura 43 - *Script*: 5-pes\_resplegend\_v0001.sql 88](#_Toc455940851)

[Figura 44 - *Script*: 0-schema e usuario.sql 88](#_Toc455940852)

[Figura 45 - *Select* na *view* “fato\_acidentes” 89](#_Toc455940853)

[Figura 46 – *Select* na *view* “fato\_uso\_alcool” 89](#_Toc455940854)

[Figura 47 – *Select* na *view* “dim\_un\_fed” 89](#_Toc455940855)

[Figura 48 – *Select* na *view* “dim\_sexo” 89](#_Toc455940856)

[Figura 49 – *Select* na *view* “dim\_fa\_eta” 90](#_Toc455940857)

[Figura 50 – *Select* na *view* “dim\_cor\_raca” 90](#_Toc455940858)

[Figura 51 – *Select* na *view* “dim\_est\_civil” 90](#_Toc455940859)

[Figura 52 – *Select* na *view* “dim\_sabe\_ler\_esc” 90](#_Toc455940860)

[Figura 53 – *Select* na *view* “dim\_curso\_que\_freq” 91](#_Toc455940861)

[Figura 54 *Select* na *view* “dim\_possui\_curso\_sup\_comp” 91](#_Toc455940862)

[Figura 55 – *Select* na *view* “dim\_dir\_car” 91](#_Toc455940863)

[Figura 56 – *Select* na *view* “dim\_dir\_mot” 91](#_Toc455940864)

[Figura 57 - *Select* na *view* “dim\_freq\_anda\_aut\_van\_taxi” 91](#_Toc455940865)

[Figura 58 - *Select* na *view* “dim\_envol\_acid\_c\_lesoes\_ult\_12m” 92](#_Toc455940866)

[Figura 59 - *Select* na *view* “dim\_ocup\_veic\_acid\_mais\_grave” 92](#_Toc455940867)

[Figura 60 - *Select* na *view* “dim\_freq\_costu\_consu\_beb\_alc” 92](#_Toc455940868)

[Figura 61 - Select na view “dim\_dirigiu\_apos\_consu\_beb\_alc” 92](#_Toc455940869)

[Figura 62 - *Select* na *view* “dim\_ult\_30d\_consu\_mais\_4\_dos\_beb\_alc\_un\_vez” 92](#_Toc455940870)

[Figura 63 – *Select* na *view* “dim\_qt\_dia\_mes\_consu\_mais\_4\_dos\_beb\_alc\_un\_vez” 93](#_Toc455940871)

Lista de Tabelas

[Tabela 1 - Comparação entre ferramentas de *Business Intelligence* 32](#_Toc455668727)

[Tabela 2 - Projetos de *Business Intelligence* desenvolvidos através do DATASUS 36](#_Toc455668728)

[Tabela 3 - Módulos da Pesquisa Nacional da Saúde 2013 38](#_Toc455668729)

[Tabela 4 - Fato “acidentes”: detalhamento dos atributos 48](#_Toc455668730)

[Tabela 5 - Fato “uso\_alcool”: detalhamento dos atributos 50](#_Toc455668731)

[Tabela 6 - Medidas com valores diferentes de "1" 60](#_Toc455668732)

[Tabela 7 - Atributos do cubo “acidentes” 61](#_Toc455668733)

[Tabela 8 - Atributos do cubo “uso\_alcool” 62](#_Toc455668734)

[Tabela 9 – Atributos do cubo “acid\_vs\_uso\_alc” 73](#_Toc455668735)

Lista de Abreviaturas e Siglas

|  |  |
| --- | --- |
| BI | *Business Intelligence* |
| BSC | *Balanced Scorecard* |
| CRM | *Customer Relationship Management* |
| DATASUS | Departamento de Informática do SUS |
| DW | *Data Warehouse* |
| ETL | *Extract Transform Load* |
| FUNASA | Fundação Nacional de Saúde |
| IBGE | Instituto Brasileiro de Geografia e Estatística |
| KDD | *Knowledge Discovery in Databases* |
| OLAP | *On-Line Analytic Processing* |
| PNAD | Pesquisa Nacional por Amostra de Domicílios |
| SGBD | Sistema de Gerenciamento de Banco de Dados |
| SUS | Sistema Único de Saúde |
| IDB | Indicadores e Dados Básicos |

Sumário

[Introdução 13](#_Toc455668736)

[*1* *Business Intelligence* 15](#_Toc455668737)

[*1.1* *Data Warehouse* 16](#_Toc455668738)

[*1.1.1* *Data Mart* 17](#_Toc455668739)

[*1.1.2* *Extract Transform Load* 17](#_Toc455668740)

[1.1.3 Granularidade 18](#_Toc455668741)

[1.1.4 Modelagem dimensional 19](#_Toc455668742)

[1.2 Análise de negócios 22](#_Toc455668743)

[1.2.1 *ON-LINE ANALYTIC PROCESSING* 22](#_Toc455668744)

[*1.2.2* *Data Mining* 24](#_Toc455668745)

[*1.3* *Balanced Scorecard* 25](#_Toc455668746)

[2 Ferramentas 27](#_Toc455668747)

[*2.1* *Pentaho Community* 27](#_Toc455668748)

[*2.2* *Oracle Business Intelligence* 28](#_Toc455668749)

[2.3 IBM Cognos 30](#_Toc455668750)

[*2.4* *Microsoft SQL Server* 31](#_Toc455668751)

[2.5 Análise das ferramentas 32](#_Toc455668752)

[3 DATASUS 34](#_Toc455668753)

[3.1 Tipos de Dados 35](#_Toc455668754)

[3.2 Ferramentas de *Business Intelligence* do DATASUS 36](#_Toc455668755)

[3.3 Definição dos dados para análises 37](#_Toc455668756)

[4 DESENVOLVIMENTO DO *DATA WAREHOUSE* E DOS PROCESSOS ETL 40](#_Toc455668757)

[4.1 Etapas da criação do *Data Warehouse* e processos ETL 40](#_Toc455668758)

[4.1.1 Extração e criação do *Data Warehouse* 41](#_Toc455668759)

[4.1.2 Transformação e Carga 42](#_Toc455668760)

[4.2 Modelagem multidimensional do *Data Warehouse* 47](#_Toc455668761)

[4.2.1 *View* “fato\_acidentes” 47](#_Toc455668762)

[4.2.2 View “fato\_uso\_alcool” 49](#_Toc455668763)

[5 IMPLEMENTAÇÃO DA FERRAMENTA DE BI 52](#_Toc455668764)

[5.1 Preparação e instalação do ambiente 52](#_Toc455668765)

[5.2 Conexão da ferramenta com o *Data Warehouse* 55](#_Toc455668766)

[*5.3* Criação e configuração dos cubos (*Data Sources)* 56](#_Toc455668767)

[5.3.1 Cubo “acidentes” 58](#_Toc455668768)

[5.3.2 Cubo “uso\_alcool” 62](#_Toc455668769)

[5.3.3 Restrição de realizar um único *join* entre a fato e uma dimensão no *Pentaho Community* 63](#_Toc455668770)

[5.4 Visualização OLAP e avaliação dos resultados 64](#_Toc455668771)

[5.4.1 Visualização OLAP do cubo “acidentes” 65](#_Toc455668772)

[5.4.2 Visualização OLAP do cubo “uso\_alcool” 69](#_Toc455668773)

[5.4.3 Criação e visualização OLAP do cubo “acid\_vs\_uso\_alc”: cruzamento dos cubos “acidentes” e “uso\_alcool” 72](#_Toc455668774)

[5.5 Entrevistas 77](#_Toc455668775)

[CONCLUSÃO 80](#_Toc455668776)

[Referências Bibliográficas 82](#_Toc455668777)

[ANEXO A - Tabela de Projetos de Business Intelligence desenvolvidos através do DATASUS 84](#_Toc455668778)

[APÊNDICE A – *SCRIPTS* PARA A CRIAÇÃO DO *DATA WAREHOUSE* 88](#_Toc455668779)

[APÊNDICE B – CONTEÚDO DAS *VIEWS* FATO E DIMENSÕES, LIMITADO A 5 REGISTROS 89](#_Toc455668780)

# Introdução

Atualmente, no mundo todo, frente aos desafios que se apresentam no ambiente corporativo, as empresas têm buscado ampliar o uso de ferramentas que facilitem a tomada de decisão baseada em informações úteis e atuais. Com os constantes e grandes avanços da tecnologia da informação, estas empresas podem contar com recursos que possibilitam armazenar, processar e controlar enormes volumes de dados. Neste contexto as ferramentas mais utilizadas são as de *Business Intelligence* (BI), que segundo Jacobson e Misner (2007), subentende-se por um conjunto de tecnologias que faz uso de informações já disponíveis em uma empresa. Estas tecnologias tem como principais objetivos, proporcionar o acesso interativo aos dados e permitir a manipulação desses dados. O processo do BI baseia-se na transformação de dados em informações, depois em decisões e finalmente em ações.

Com o intuito de adquirir informações para melhorar a gestão de seus negócios, as empresas podem utilizar de técnicas e ferramentas para integrar este grande volume de dados. Através de processos de *Extract Transform Load* (ETL), estas informações, que podem ser provenientes de diversas fontes de dados, são integradas dentro do *Data Warehouse* para que possam ser utilizados por ferramentas voltadas para o acesso com diferentes níveis e formas de apresentação.

Em muitos casos os sistemas utilizados dentro de uma organização não atendem de forma adequada às necessidades de seus gestores, não mostrando as informações necessárias para que possam tomar decisões. É necessário que estes gestores tenham em mãos ferramentas simples, ágeis e capazes de gerar informações que possam auxilia-los nos processos de tomada de decisão.

No entanto, apesar da área da saúde ser considerada um setor vital, o uso destas ferramentas para tomadas de decisões ainda não está consolidado. No Brasil todos os dados informatizados referentes ao Sistema Único de Saúde (SUS), estão disponíveis no Departamento de Informática do SUS (DATASUS), que trata-se de um órgão que desenvolve pesquisas e incorpora tecnologias de informática que possibilitam a implementação de sistemas e a disseminação de informações necessárias as ações de saúde.

Utilizando um ambiente de *Business Intelligence*, este estudo analisou fontes de dados abertas, que foram extraídas de órgãos públicos do sistema de saúde brasileiro com a ajuda ferramentas e técnicas de *ETL.* Com estes dados, foram desenvolvidos cubos e visões, utilizando técnicas e conceitos *On-Line Analytic Processing* (OLAP*)*, objetivando a extração de informações relevantes e de interesse social a respeito da saúde pública no Brasil.

Este trabalho está dividido em cinco capítulos. No primeiro capítulo foi abordado o conceito de *Business Intelligence* e seus componentes, dando um enfoque maior nos termos *Data Warehouse* e Análise de negócios. No capítulo seguinte foi realizada uma pesquisa com algumas ferramentas de *Business Intelligence*, das quais uma foi escolhida para a elaboração deste trabalho. Já no terceiro capítulo foram abordados os dados informatizados referentes ao Sistema Único de Saúde (SUS), os quais encontram-se disponíveis no Departamento de Informática do SUS (DATASUS). Além disso no terceiro capítulo, foram definidos quais os dados para realizar as análises pela ferramenta de BI. No capítulo quatro foi realizada a criação e modelagem do *Data Warehouse*, além das etapas e processos de ETL. No quinto e último capítulo, implementou-se a ferramenta de BI, onde foi possível analisar os resultados obtidos.

# *Business Intelligence*

Este capítulo tem como objetivo apresentar os conceitos, técnicas e ferramentas de *Business Intelligence* e como seu uso pode ajudar nos processos de tomadas de decisões dentro de um ambiente corporativo. Nele são destacados os principais temas que compõem um ambiente de *Business Intelligence*: *Data Warehouse*, *Data Marts*, *ETL*, Granularidade, Modelagem dimensional, *Online Analytical Processing*, *Data Mining* e *Balanced Scorecard*.

O conceito de *Business Intelligence* já existe há milhares de anos. As antigas civilizações, assim como empresas e grandes organizações nos dias atuais, necessitam tomar decisões. Na antiguidade, as civilizações baseavam-se em dados e informações provenientes da natureza, como por exemplo, o comportamento dos mares e os períodos de chuva, para então, com base nestes dados e informações, tomarem suas decisões para melhorar e atender seus interesses da época.

Na atualidade, de uma forma mais abrangente, conforme Barbieri (2001), *Business Intelligence* pode ser entendido como a utilização de variadas fontes de informação para definir estratégias nos negócios de uma empresa. É um processo inteligente de coleta, organização, análise, comportamento e monitoramento de dados contidos dentro de um *Data Warehouse* e seus *Data Marts*.

O processo de *Business Intelligence* baseia-se na transformação dos dados em informação, depois em decisões e finalmente em ações. Em outras palavras, é um conjunto de ferramentas e técnicas que objetivam dar suporte a tomada de decisão. O processo resume-se em extrair dados, analisá-los de forma eficiente, para que sejam transformados em informações.

Dentro das empresas e grandes organizações, o atual interesse pelo *Business Intelligence* vem crescendo, na medida em que sua utilização possibilita agilizar os processos relacionados às tomadas de decisões. A necessidade de se manterem em um mercado altamente competitivo e mutável faz com que elas busquem novas formas para se diferenciarem. O *Business Intelligence* é a uma poderosa ferramenta que elas podem utilizar para o alcance do sucesso e obter uma vantagem competitiva frente aos seus concorrentes.

O uso de ferramentas de *Business Intelligence* possibilita que gestores tenham mais informações e condições de tomar decisões certas, sem agir por impulso ou por palpites. É necessário que eles tenham em mãos informações mais eficientes, que mostrem uma real necessidade da empresa para que possam realizar uma série de análises e projeções.

Em sua grande, maioria as empresas possuem diversas fontes de dados, porém existe uma grande dificuldade na extração de informações a partir deles. Os diversos sistemas dentro de uma organização não trazem as informações gerenciais na sua forma mais objetiva. Estas informações necessárias para as tomadas de decisões ficam escondidas em inúmeras tabelas e arquivos de sistemas. Em contrapartida, os dados contidos nestes sistemas possuem um grande potencial para serem analisados, desde que passem pelo processo de extração, transformação e armazenamento dentro de um repositório de dados, para então ser feita a exploração dos dados, através de ferramentas de *Business Intelligence*.

Neste contexto, conforme afirma Barbieri (2001), o principal objetivo das ferramentas de *Business Intelligence*, como por exemplo, *Data Warehouse*, *Data Mart* e ETL, está na definição de regras e técnicas para a formatação correta de todas as fontes de dados, com a finalidade de transformá-los em depósitos organizados de informação independente de sua origem.

## *Data Warehouse*

Caracteriza-se como um grande repositório de dados, os quais ficam organizados e acessíveis com as informações necessárias para gestão de negócios. Seu principal objetivo é fornecer informações confiáveis e úteis para ajudar no processo de tomada de decisão. Os dados contidos em um *Data Warehouse* podem ser provenientes de diversas fontes e sistemas existentes dentro de uma organização. Seus dados são somente inseridos, não existindo atualização (MACHADO, 2000).

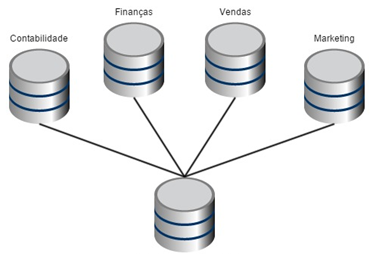
Um *Data Warehouse* possibilita uma sólida e concisa integração dos dados de uma empresa para a realização de análises gerenciais estratégicas. Ele disponibiliza acesso aos dados confiáveis, que já passaram por um processo de validação. Através de técnicas e processos de *Extraction, Transformation, Load* (ETL), os dados provenientes de diferentes fontes e sistemas em uso, dentro de uma organização, são integrados dentro do *Data Warehouse,* para que possam ser utilizados por ferramentas voltadas para o acesso com diferentes níveis e formas de apresentação.

Para desenvolver um projeto de *Data Warehouse* é necessário realizar um levantamento e descobrir as necessidades da informação, indicadores e também o entendimento do negócio da empresa. Com base nesta análise, deve-se projetar e modelar o banco de dados para armazenar os dados, que através de processos de *ETL* serão extraídos das diversas fontes analisadas para disponibilizar informações para os sistemas de *Business Intelligence*.

### *Data Mart*

Segundo Inmon, Terdeman e Imhoff (2001), os *Data Marts* são estruturas moldadas pelos dados setoriais encontrados no *Data Warehouse*. Conforme mostra a Figura 1, os *Data Marts* pertencem aos departamentos específicos dentro de uma empresa, como por exemplo, contabilidade, finanças, vendas ou marketing e são modelados de acordo com as requisições dos departamentos.

Figura - Estrutura de um *Data Warehouse* e seus *Data Marts*

  
Fonte: Adaptada de INMON, TERDEMAN E IMHOFF (2001).

Os dados armazenados dentro dos *Data Marts,* em sua grande maioria, são de uso exclusivo de um departamento ou uma área específica do negócio e representam um subconjunto de dados do *Data Warehouse*. Eles normalmente são moldados em um esquema estrela, com a finalidade de atender as necessidades de cada setor. Sua principal vantagem é a possibilidade de respostas rápidas e precisas para o usuário final (MACHADO, 2000).

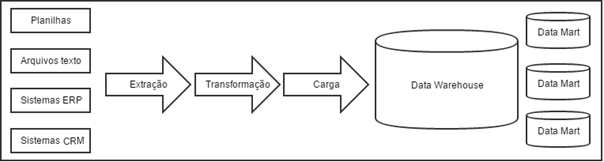
### *Extract Transform Load*

O objetivo dos processos de Extract Transform Load (ETL) é fazer a integração de informações de diversas fontes de dados. Estes processos têm a capacidade de se conectarem com diversas fontes de dados, ler os dados, transformar os dados e carregá-los em um *Data Warehouse e/*ou *Data Marts*. As informações carregadas podem ser combinadas e provenientes das mais variadas fontes de dados, como por exemplo: sistemas *CRM* (*Customer Relationship Management*), planilhas ou arquivos texto.

Para dar início a este processo, primeiramente é necessário definir as origens das fontes de dados e fazer a extração delas. Tendo em vista que os dados se originam de diferentes fontes, às vezes é necessário padronizar os diferentes formatos encontrados. Um exemplo de padronização poderia ser o uso dos valores “H” ou “M” para um atributo. Neste caso, ambos valores estariam se referindo ao sexo de uma pessoa, “H” para homem e “M” para masculino. No entanto, utiliza valores distintos para referir-se a um mesmo significado. Portanto ao fazer o processo de ETL, os dados são transformados para um formato uniforme, normalmente sugerido pelo próprio usuário para que possam ser carregados no repositório de dados.

A carga dos dados consiste em fisicamente, seguindo um modelo dimensional, estruturar e carregar os dados para dentro do *Data Warehouse*. De forma geral o processo pode ser entendido conforme demonstra a Figura 2.

Figura - Processo de *ETL*

Fonte: Autor 2015.

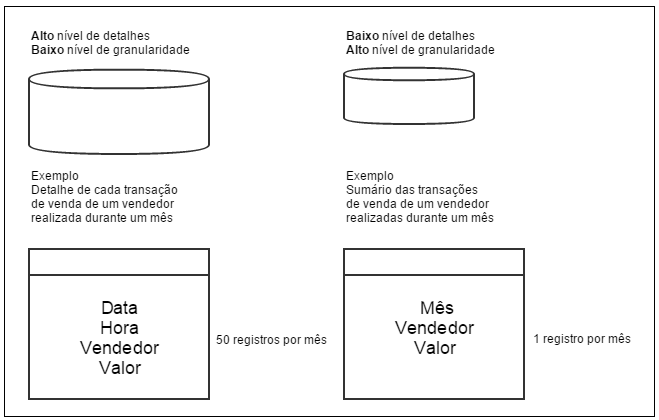
### Granularidade

A granularidade refere-se ao nível de detalhamento dos elementos disponíveis nos dados dentro de um *Data Warehouse*. Quanto menor o nível de detalhe existir nos dados, maior será a granularidade. Consequentemente, quanto maior o nível de detalhe existir nos dados, menor será a granularidade. A seleção dos níveis de granularidade terá um alto impacto no volume de dados de um *Data Warehouse* e também servirá como um parâmetro de avaliações de performance.

O volume de dados poderá ser estimado através de uma média de combinações entre as dimensões. Caso seja necessário armazenar todas as vendas de um período de 2 anos para uma rede de 200 lojas, onde cada loja realiza em média a venda de 10.000 produtos por dia, o volume estimado poderá ser calculado realizando a seguinte equação: (média de vendas por dia \* quantidade de lojas \* quantidade de dias que se deseja armazenar) ou (10.000 \* 200 \* 730), o resultado seria em torno de 1,46 bilhão de registros (MACHADO, 2000).

Conforme mostra a Figura 3, para o elemento vendas, pode-se optar por dois tipos de níveis de granularidade, de acordo com a necessidade do usuário final. Um nível de granularidade mostra a sumarização de todas as vendas agrupadas mês a mês e o outro nível mostra as vendas a cada operação realizada.

Figura - Granularidades diferentes em um mesmo assunto: Vendas



Fonte: Adaptada de Machado (2000).

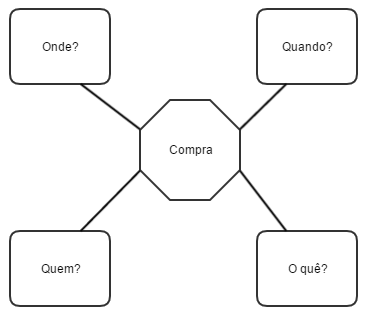
A escolha correta do nível ou níveis de granularidade a ser definido é extremamente importante para obter sucesso em um projeto, pois dependendo do nível de detalhe escolhido, as informações poderão ou não atender as necessidades do usuário final. Conforme consta na Figura 3, caso seja definida uma granularidade maior, mesmo que o usuário final precise, não será possível obter uma análise de todas as vendas realizadas durante um mês de um determinado vendedor.

### Modelagem dimensional

A modelagem dimensional é o nome dado a técnica de um projeto de *Data Warehouse*, onde as informações se relacionam de forma que podem ser representadas como um cubo. Sendo assim é possível dividir este cubo e navegar em cada uma de suas dimensões ou eixos. Basicamente esta modelagem pode ser implementada através de dois modelos, *Star Schema* e *Snowflake Schema*. Em ambos os modelos, existe uma tabela dominante no centro com múltiplas junções a outras tabelas.

De acordo com Machado (2000), dimensões são os elementos e as entidades que participam de algum fato, conforme pode-se observar na Figura 4. Por exemplo, toda ocorrência de uma compra envolve no mínimo quatro elementos: quando foi realizada a compra, onde foi realizada a compra, quem realizou a compra e o que foi comprado.

Figura - Fato compra e seus elementos

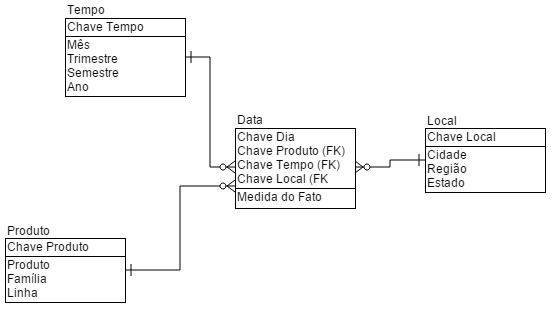
  
Fonte: Adaptado de Machado (2000).

Independente do modelo utilizado, para realizar esta modelagem existem dois tipos de tabelas, são elas:

* Tabela Fato: a tabela dominante, que fica no centro do modelo, é responsável por armazenar as medidas numéricas referentes a situação de negócio. Cada uma das medidas é obtida a partir da união de todas as dimensões. Ela também armazena as chaves de junções para as tabelas de dimensões.
* Tabela Dimensão: é conectada a tabela fato. Nela ficam armazenadas as descrições textuais das dimensões, fornecendo um maior nível de detalhamento para o negócio. Seu relacionamento é de um para muitos, o que possibilita um melhor aproveitamento do processamento das informações para as tabelas fato. Possuem diversos atributos de todos os tipos de dados, podendo armazenar as mais variadas informações (KIMBALL; ROSS, 2013).

O modelo estrela (*Star Schema*) é a estrutura básica de um modelo de dados multidimensional. Ele é composto por uma tabela central denominada fato e um conjunto de tabelas menores denominadas dimensões, as quais ficam dispostas ao redor desta tabela central formando uma estrela (MACHADO, 2000).

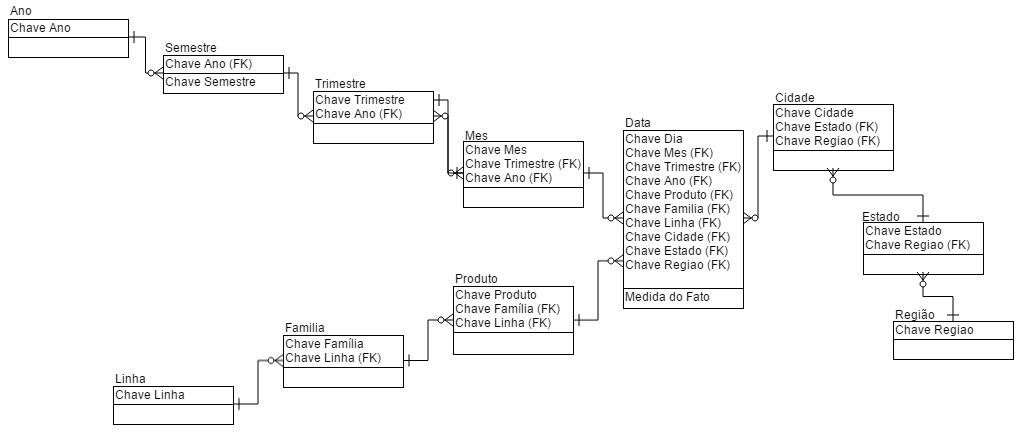
Figura - Estrutura de um Modelo Estrela (*Star schema*)

  
Fonte: Adaptado de Machado (2000).

Conforme pode-se observar na Figura 5, o centro da estrela é o fato e ao seu redor estão dispostas as dimensões que participam deste fato. O relacionamento entre a tabela central e as tabelas dimensões é uma simples ligação entre as entidades em um relacionamento de um para muitos.

Conforme Machado (2000), o modelo *SnowFlake* (Floco de Neve) tem dimensões separadas em hierarquias, uma tabela para cada nível. A hierarquia é o resultado da decomposição de uma ou mais dimensões que possuem hierarquia entre seus membros, ou seja relacionamentos muitos para um entre os membros de uma dimensão, formando relacionamentos entre tabelas dimensões. Sua utilização preserva os meios de armazenamento, como ocorre em um modelo normalizado, desta forma, ele evita a redundância de valores textuais em uma tabela.

Figura - Estrutura de um Modelo Floco de Neve (*Snowflake*)



Fonte: Adaptado de Machado (2000).

Na Figura 6 é possível notar que o relacionamento entre a tabela fato e as tabelas dimensões ocorre seguindo uma hierarquia, com várias tabelas de sub dimensões, onde cada tabela é responsável por um nível de detalhe.

Conforme pode-se observar nos dois modelos, *Star Schema* e *SnowFlake*, ambos possuem uma tabela central, denominada fato e disposta ao seu redor as tabelas dimensões. É possível notar uma grande semelhança entre ambos, porém no modelo *Star Schema* não ocorre a normalização dos dados contidos nas tabelas dimensões ao contrário do que ocorre no modelo *SnowFlake*, onde as tabelas fatos são normalizadas e possuem hierarquia entre suas tabelas.

## Análise de negócios

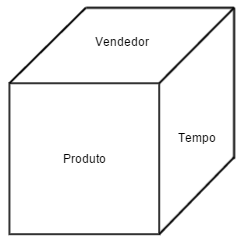
Análise de negócios é um amplo conjunto de ferramentas e técnicas para reunir, armazenar, analisar e fornecer informações, com o objetivo de ajudar os gestores de uma organização a tomarem as melhores decisões. Seu uso emprega ferramentas *OLAP*, *Data Mining* e técnicas de análises de dados multidimensionais.

### *ON-LINE ANALYTIC PROCESSING*

É um conjunto de ferramentas capazes de visualizar e explorar os dados armazenados dentro de um *Data Warehouse*. Estes dados são exibidos usando uma análise denominada multidimensional, representando os dados em dimensões e não em tabelas. Combinando estas dimensões o usuário pode criar diversas visualizações, combinando entre os níveis de detalhamento (MACHADO, 2000).

*OLAP* é uma ferramenta que complementa o *Data Warehouse*. Enquanto o *Data Warehouse* armazena os dados, a *OLAP* os recupera com maior eficiência e velocidade, transformando estes dados em informações. Isto ocorre de tal forma que o usuário possa visualizar e ter conhecimento destes dados, e possa analisá-los com profundidade e explorá-los em diversas dimensões e perspectivas, representando os dados em forma de um cubo, conforme demonstra a Figura 7.

Figura - Cubo *OLAP*: Vendas



Fonte: Autor, 2015.

*OLAP* é considerado uma ferramenta de apoio à decisão que permite análise multidimensional de dados do *Data Warehouse* e seus *Data Marts*. Seu uso auxilia a empresa na elaboração de estratégias e também na tomada de decisões, proporciona aos usuários finais condições para extrair os dados de suas bases e construir relatórios capazes de responder às suas questões gerenciais. A análise multidimensional proporciona a flexibilidade analítica para responder perguntas como:

* Como o gasto com propaganda afetou as vendas?
* Onde os concorrentes estão fazendo incursões?
* Quais produtos estão perdendo participação de mercado?
* Quais são os clientes mais fiéis?

Estas respostas exigem um conjunto de técnicas capazes de extraí-las dos dados armazenados no *Data Warehouse*. Através destas técnicas os usuários finais serão capazes de gerar consultas e relatórios, análises multidimensionais, análises estatísticas e também *Data Mining*.

Além das quatro primeiras operações, citadas por Machado (2000), também existem outras duas, conforme Barbieri (2001), são elas:

* *Drill Down e Roll Up:* onde *Drill Down* aumentar um nível de detalhe dentro de uma visualização e *Roll Up*, diminui um nível de detalhe em uma visualização.
* *Drill Across*: ocorre quando um usuário pula um nível de detalhe dentro de uma visualização.
* *Drill Throught*: permite uma busca de informações além do nível de granularidade existente na estrutura dimensional.
* *Slice and Dice*: com esta operação o usuário pode reduzir o escopo dos dados em análise, além de mudar a ordem das dimensões e orientação de como os dados são visualizados.
* *Ranking*: classifica determinada informação, filtrando os dados apresentados para o usuário, ajudando no processo de tomada de decisão.
* *Pivoting*: significa analisar a mesma informação em diversos ângulos. Por exemplo alternar a visualização entre linhas e colunas.

Além das operações de OLAP que podem ser combinadas e utilizadas ao mesmo tempo, é possível modificar as posições das dimensões e alternar entre linhas em colunas dentro de uma planilha. Este recurso é chamado de *pivot*, que é a capacidade de visualizar os dados em diversos ângulos.

### *Data Mining*

*Data Mining* é uma das etapas do processo de descoberta de conhecimento em banco de dados, também conhecido como *KDD* (*Knowledge Discovery in Databases*). O processo de *KDD* é composto por seis fases, que envolvem a seleção dos dados, limpeza, enriquecimento, transformação, *Data Mining* e construção de relatórios para apresentação das informações descobertas (ELMASRI; NAVATHE, 2011).

Como a tradução do nome indica, *Data Mining* refere-se à extração ou descoberta de novas informações e conhecimento armazenados em bancos de dados. É o processo de descobrir informações relevantes em grandes volumes de dados, como padrões, associações, mudanças, anomalias e estruturas. Pode ser usado em conjunto com um *Data Warehouse* para auxiliar a encontrar novas informações que não podem ser obtidas através de consultas. Em geral o processo de *Data Mining* divide-se em:

* Preparação: é onde os dados são preparados para serem disponibilizados às técnicas de *Data Mining*. Nesta fase os dados mais importantes são selecionados e são retiradas as suas inconsistências.
* *Data Mining*: nesta fase os dados preparados são processados e transformados para uma forma onde possam ser identificadas as informações mais importantes.
* Análise de Dados: é a etapa onde resultado do *Data Mining* é avaliado, buscando determinar se algum conhecimento foi descoberto, assim como definir a importância dos fatos gerados.

A apresentação dos resultados do processo de *Data Mining* pode ser exibida em diversos formatos como por exemplo: listagens, planilhas resumidas ou gráficos (ELMASRI; NAVATHE, 2011).

## *Balanced Scorecard*

O objetivo de *Balanced Scorecard* (*BSC*) é medir e gerir o desempenho de uma empresa, alinhando o seu planejamento estratégico às suas atividades operacionais. Barbieri (2001) utiliza o conceito de balanceamento de quatro perspectivas fundamentais para a avaliação correta da empresa e de seus projetos. São eles:

* O retorno dos recursos financeiros neles investidos;
* A satisfação dos clientes que consomem seus produtos e serviços;
* A agilidade, fluidez e melhoria dos processos internos que os executam e controlam;
* A dimensão do recurso *peopleware*, com seus anseios, frustrações, evoluções e alinhamentos.

De acordo com seus criadores Kaplan e Norton (2006), a capacitação das pessoas promove uma melhoria nos processos da empresa que, por sua vez, satisfazem mais aos clientes, proporcionando melhores resultados financeiros. Ferramentas de *Business Intelligence* estão diretamente ligadas aos projetos que envolvem *BSC*. Através delas são fornecidos dados históricos e indicadores que são extraídos dos *Data Warehouse* e *Data Marts* das áreas funcionais de uma empresa.

Atualmente, a maioria das ferramentas de *Business Intelligence* permitem a aplicação de *BSC* através da capacidade de comparar imediatamente o desempenho de negócios com as metas estabelecidas, permitindo que gestores rapidamente compreendam como vai o negócio da empresa (TURBAN et al., 2009).

# Ferramentas

Neste capítulo serão mostradas algumas ferramentas de *Business Intelligence* disponíveis no mercado, será analisado e exemplificado o uso de cada uma, buscando utilizar e aplicar os conceitos descritos no capitulo 1. Após esta análise será feita a definição da ferramenta que será utilizada para o desenvolvimento deste projeto. As ferramentas estudadas são o *Pentaho Community*, *IBM Cognos*, *Oracle Business Intelligence Foundation* e *Microsoft SQL Server*.

## *Pentaho Community*

*Pentaho Community* é um conjunto de componentes *open source* para *Business Intelligence*. Foi desenvolvido em Java no ano de 2004 pela empresa *Pentaho Corporation* nos Estados Unidos. A solução fornece um amplo e completo conjunto de ferramentas para análise de negócio, como por exemplo: *OLAP*, *ETL*, relatórios e *Data Mining* (HITACHI, 2015).

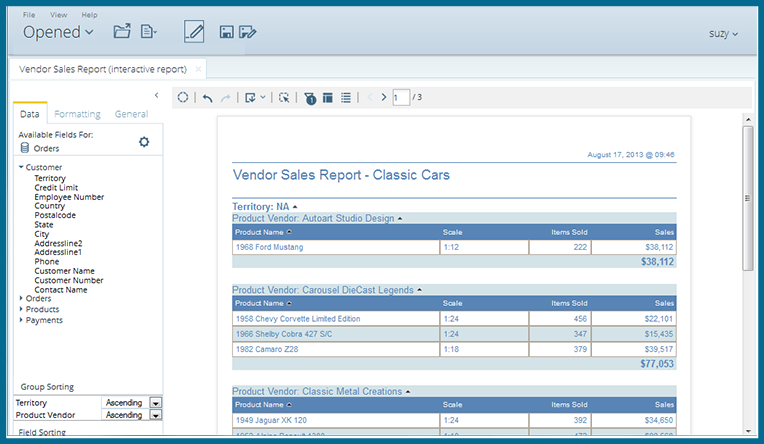
Por ser uma plataforma de código aberto, desenvolvedores podem montar rapidamente soluções personalizadas de BI, aumentando o valor e a capacidade das soluções de suas empresas, criando propostas superiores as tradicionais do mercado a um custo significativamente baixo.

O conjunto de ferramentas do *Pentaho Community* possibilita aos usuários, extrair, transformar, carregar, acessar, descobrir e mesclar qualquer tipo de dados, independentemente da sua fonte e dimensão, através dos seguintes módulos:

* *OLAP* – *jPivot*: oferece acesso interativo aos dados, permitindo, cruzar, realizar consultas e análises multidimensionais em um grande volume de dados.
* *ETL*- *Kettle*: é o modulo responsável pela integração de dados do *Pentaho* através de diversas fontes.
* Relatórios – *Report Designer*: permite desenvolver relatórios relacionais e analíticos de uma ampla fonte de dados.
* *Data Mining* – *Weka*: é um abrangente conjunto de ferramentas e técnicas para mineração de dados. Seu uso possibilita descobrir padrões e extrair informação em grandes conjuntos de dados.

A Figura 8 demonstra um exemplo de relatório interativo desenvolvido utilizando o *Pentaho Report Designer*.

Figura - Relatórios interativos no Pentaho



Fonte: Hitachi, 2015.

Os seus módulos podem ser instalados de forma independentes de acordo com a necessidade, oferecendo flexibilidade, extensibilidade e escalabilidade para atender as necessidades de TI, dentro de um ambiente totalmente unificado. Seu uso promove diversas habilidades, agregando valor as estratégias de negócios dentro de uma empresa.

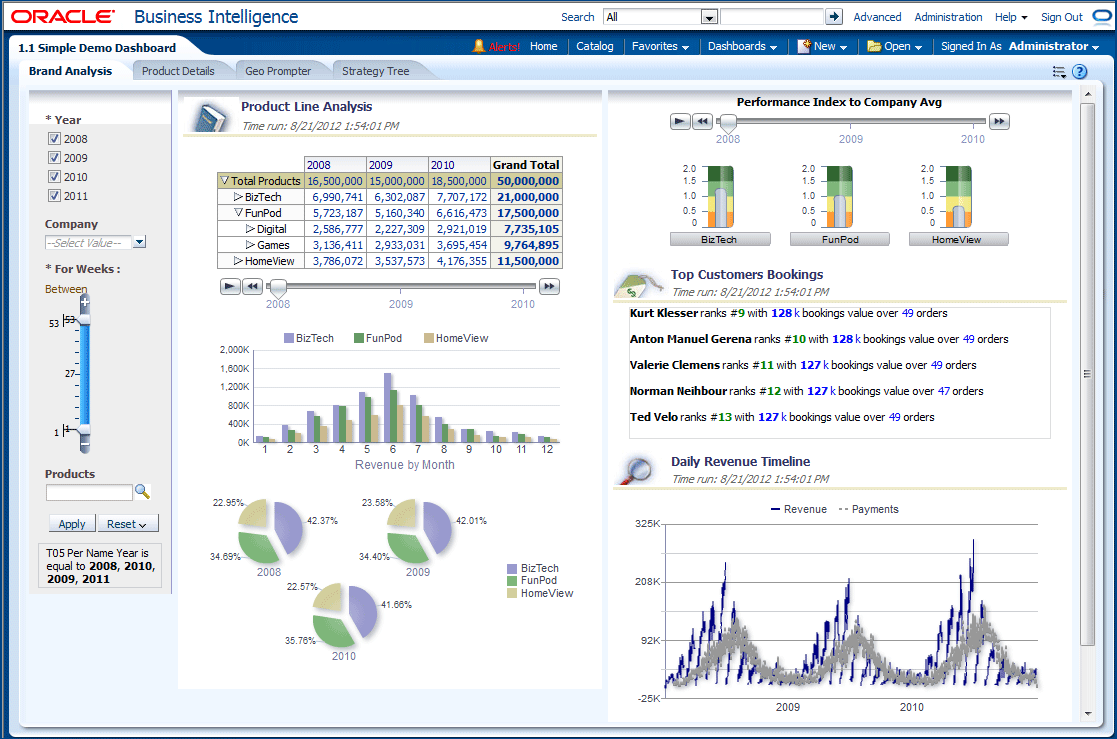
## *Oracle Business Intelligence*

O *Oracle Business Intelligence* é uma suíte completa de componentes de *Business Intelligence* oferecida pela *Oracle Corporation*, o qual disponibiliza um conjunto bem diversificado de relatórios, consultas, painéis interativos, gráficos animados e análises *OLAP*. Seu uso oferece à empresa uma visão abrangente do negócio em um só lugar, ajudando a organização a impulsionar a inovação e otimizar as operações (ORACLE, 2015).

Dentre todo conjunto de componentes que compõem a suíte, pode-se destacar os principais:

* *Oracle Exalytics In-Memory Engineered System*: é um sistema projetado especificamente para fornecer análises de alta performance em tempo real.
* *Oracle Business Analytics*: é o componente central do conjunto de componentes para a criação de relatórios, consultas, painéis interativos, gráficos animados e análises *OLAP*.
* *Oracle Scorecard and Strategy Management*: seu uso ajuda as empresas a estabelecer metas e objetivos estratégicos, é ligado a indicadores de desempenho para que possam ser medidos, rastreados e gerenciados.
* *Oracle Business Intelligence Publisher*: é a solução para a publicação de relatórios em vários formatos. Os usuários podem criar e modificar interativamente e publicar os relatórios utilizando um navegador web.
* *Oracle Business Intelligence Mobile*: dispõe aos usuários visualizar e interagir com seus dados de forma otimizada, utilizando dispositivos móveis.
* *Oracle Warehouse Builder*: é uma solução oferecida para realizar a extração e integração dos dados.
* *Office Integration*: permite que seus usuários importem análises e *dashboards* em seus documentos do *Office*, bem como criar análises diretamente do Excel e guardá-las para o catálogo do *Oracle BI*. A Figura 9 demonstra uma *dashboard* exibindo tabelas, gráficos e analises dimensionais.

Figura - Oracle *Business Intelligence* exemplo de uma *dashboard*



Fonte: Oracle, 2015.

Conforme pode-se observar o *Oracle Business Intelligence* oferece uma ampla gama de componentes com capacidades de análise e relatórios em um único lugar, utilizando uma tecnologia totalmente completa e integrada para análises de negócios (ORACLE, 2015).

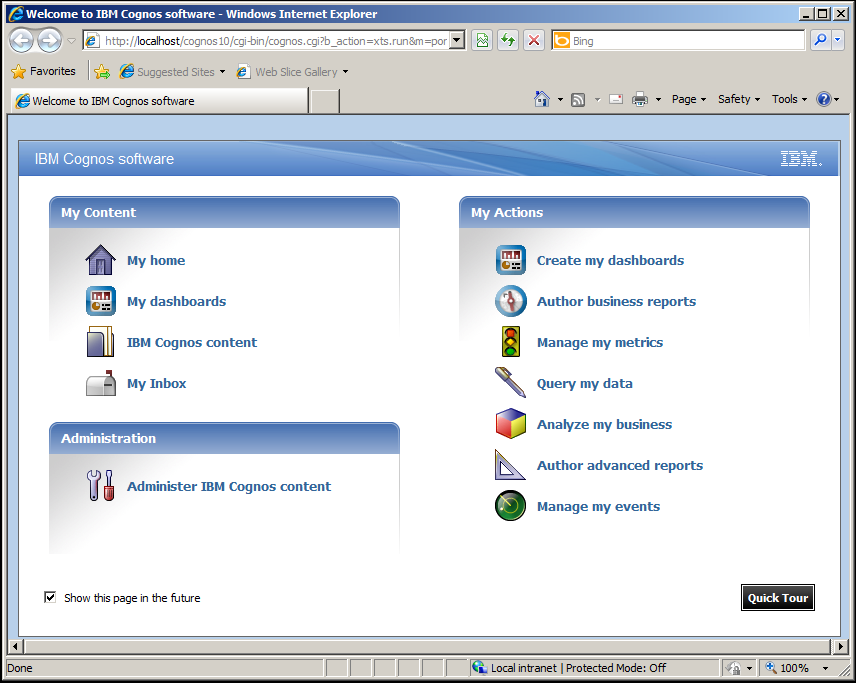
## IBM Cognos

IBM Cognos é uma solução integrada para planejamento e inteligência de negócios, *Business Intelligence,* desenvolvida pela *International Business Machines* (IBM). Através dela é possível criar consultas, relatórios, análises, painéis e indicadores que podem ser estendidos com planejamentos, modelagens de cenários e análises de previsão (IBM, 2015). Ela ajuda no processo de tomada de decisão baseado em interações, pesquisa e montagem das perspectivas de um negócio. Dentre todas ferramentas oferecidas por todo conjunto, pode-se destacar:

* IBM *Cognos Query Studio*: seu objetivo é permitir ao usuário realizar consultas simples para resolver rapidamente problemas específicos que possam surgir.
* *Report Studio IBM Cognos*: é a principal ferramenta para a criação de relatórios e geração de gráficos.
* *Cognos Analysis Studio*: permite a navegação em dados organizados em estruturas multidimensionais OLAP, dispõe ao usuário realizar análises mais complexas, criar comparativos e descobrir tendências a riscos.
* *Cognos Studio Metric*: é utilizado para a construção de métricas e *scorecards*, permite definir indicadores de desempenho, através do uso de diferentes fontes de dados que poderão ser exibidos em painéis, gráficos e relatórios.
* *Cognos Data Manager*: integra dados entre diversas fontes usando uma estrutura independente. Suporta a extração, transformação e carga de grandes volumes de dados.

A Figura 10 exibe a tela inicial da interface com o usuário final, que pode ser acessada através de qualquer navegador web.

Figura – Tela inicial da interface do usuário final do IBM Cognos



Fonte: IBM, 2015.

As ferramentas apresentadas no conjunto *IBM Cognos* oferecem uma solução completa e integrada para inteligência de negócios em um nível pessoal, departamental, ou de toda a empresa, ajudando as organizações a tomar as decisões que permitam atingir melhores resultados para seus negócios.

## *Microsoft SQL Server*

O *Microsoft SQL Server* é um sistema gerenciador de banco de dados desenvolvido pela *Microsoft Corporation*. Seu uso oferece uma plataforma ampla e abrangente, capaz de implantar soluções de *Business Intelligence*, integrando diversas funcionalidade e ferramentas para criar análises dos dados armazenados dentro do seu repositório de dados (MICROSOFT, 2015). Dentro do contexto deste trabalho, pode-se destacar algumas destas ferramentas:

* *SQL Server Analysis Services*: é uma ferramenta que fornece recursos *OLAP* e também funcionalidades para *Data Mining* em aplicações de *Business Intelligence*. Com o seu uso é possível projetar, criar e gerenciar estruturas multidimensionais de dados, permitindo desenvolver relatórios de negócios de forma simples e interativa.
* *SQL Server Integration Services*: é uma plataforma oferecida em conjunto com o *Microsoft SQL Server* que possibilita a extração, transformação, carga e integração de diversos sistemas e banco de dados. Através de seu uso é possível gerenciar múltiplas conexões, encaminhamento de e-mails, envios ou requisição de arquivos, permite também o envio dos dados extraídos para vários destinos, por exemplo, uma replicação de dados em um banco de dados *backup*.
* *SQL Server Reporting Services*: é um amplo ambiente de ferramentas e serviços prontos para uso, através delas é possível criar, implantar e gerenciar relatórios, painéis, planilhas e gráficos a partir do banco de dados *Microsoft SQL Server*.

Dentre todo o conjunto de ferramentas avaliadas e destacadas nesta seção, é possível observar que a Microsoft disponibiliza de forma totalmente integrada ao seu sistema de banco de dados *SQL Server*, um amplo e completo conjunto de ferramentas para *Business Intelligence*, permitindo desde a extração, transformação, carga, armazenagem até a geração de análises multidimensionais dos dados armazenados.

## Análise das ferramentas

Dentre todas as ferramentas estudadas, é possível observar e concluir que todas dispõem de excelentes conjuntos e técnicas que atendem todos os requisitos e necessidades para a elaboração deste trabalho, conforme pode-se observar na Tabela 1.

Tabela - Comparação entre ferramentas de *Business Intelligence*

| Recurso | Oracle BI | IBM Cognos | Pentaho *Community* | SQL *Server* |
| --- | --- | --- | --- | --- |
| OLAP | Oracle *Business Analytics* | Cognos *Analysis Studio* | *jPivot* | *Analysis Services* |
| *Data Mining* | Oracle Data *Miner* | Não | Weka | *Analysis Services* |
| BSC | Oracle *Scorecard and Strategy Management* | Cognos *Studio* *Metric* | *Report Designer* | *Analysis Services* |
| DW e ETL | Oracle *Warehouse Builder* | Cognos *Data* *Manager* | Kettle | *Integration Services* |
| Custo | Comercial | Comercial | Gratuita | Comercial |

Fonte: Autor, 2015.

Para a implementação da solução deste trabalho, optou-se pela ferramenta *Pentaho Community* e seus componentes. Em ordem de importância, os principais motivos por esta escolha: toda documentação disponível em seu portal de forma clara e objetiva, ser oferecida de forma gratuita e possuir uma ampla comunidade de utilizadores e desenvolvedores ao redor de todo o mundo.

# DATASUS

O Sistema Único de Saúde (SUS) é considerado um dos maiores sistemas públicos de saúde do mundo. Ele abrange desde o atendimento ambulatorial até o transplante de órgãos, garantindo acesso integral, universal e gratuito para toda a população do país. Foi criado em 1988 pela atual Constituição Federal Brasileira, em seu artigo 192, onde consta que saúde é um direito de todos e dever do estado.

O Departamento de Informática do SUS (DATASUS) trata-se de um órgão que desenvolve pesquisas e incorpora tecnologias de informática que possibilitam a implementação de sistemas e a disseminação de informações necessárias as ações de saúde. Surgiu em 1991 com a criação da Fundação Nacional de Saúde (FUNASA). Na época, a fundação passou a exercer a função de controle e processamento das contas referente a saúde que antes era da empresa de tecnologia e informações da previdência social. Então foi formalizada a criação e as competências do DATASUS. A ele compete:

I - fomentar, regulamentar e avaliar as ações de informatização do SUS, direcionadas à manutenção e ao desenvolvimento do sistema de informações em saúde e dos sistemas internos de gestão do Ministério da Saúde;

II - desenvolver, pesquisar e incorporar produtos e serviços de tecnologia da informação que possibilitem a implementação de sistemas e a disseminação de informações necessárias às ações de saúde, em consonância com as diretrizes da Política Nacional de Saúde;

III - manter o acervo das bases de dados necessários ao sistema de informações em saúde e aos sistemas internos de gestão institucional;

IV - assegurar aos gestores do SUS e aos órgãos congêneres o acesso aos serviços de tecnologia da informação e bases de dados mantidos pelo Ministério da Saúde;

V - definir programas de cooperação tecnológica com entidades de pesquisa e ensino para prospecção e transferência de tecnologia e metodologia no segmento de tecnologia da informação em saúde;

VI - apoiar os Estados, os Municípios e o Distrito Federal na informatização das atividades do SUS;

VI - apoiar os Estados, os Municípios e o Distrito Federal na informatização das atividades do SUS (Art 35 do Decreto Nº 7.530).

Observando as competências acima, é possível identificar que este trabalho tem relação com as competências II e III. Através delas será desenvolvido um ambiente de BI, utilizando o acervo de dados disponíveis no DATASUS, o qual é apresentado no decorrer deste trabalho.

## Tipos de Dados

Em seu portal através do aplicativo TABNET, o DATASUS disponibiliza dados que podem servir para auxiliar análises objetivas da situação sanitária, além de tomadas de decisão baseadas em evidências e elaboração de programas de ações de saúde.

O TABNET é um aplicativo desenvolvido pelo DATASUS que está disponível a toda a população brasileira. Sua função é gerar informações a partir das bases de dados do SUS. De acordo com a informação que se deseja obter, é possível montar consultas de forma rápida e flexível, exibindo os dados que ficam organizados em linhas e colunas (DATASUS A, 2015).

Através do aplicativo TABNET é possível gerar tabelas bidimensionais e obter informações, tais como:

* Indicadores de Saúde: Indicadores e Dados Básicos (IDB) / Cadernos de Saúde / Pactuação
* Assistência à Saúde: Sistemas Hospitalares / Ambulatoriais / Imunizações / Atenção Básica / Vig.Nutricional
* Epidemiológicas e Morbidade: Sistemas de Morbidade / Agravos / Nutrição
* Rede Assistencial: Cadastro Nacional Estabelecimento Saúde
* Estatísticas Vitais: Sistema de Mortalidade / Nascidos Vivos
* Demográficas e socioeconômicas: População / PIB / Saneamento
* Inquéritos e Pesquisas: Pesquisa Nacional por Amostra de Domicílios (PNAD). VIGITEL / VIVA
* Saúde Suplementar: Agência Nacional de Saúde Suplementar
* Estatísticas de acesso ao TABNET: estatísticas de acesso.

O TABNET teve seu início com o registro sistemático de dados de mortalidade e de sobrevivência (Estatísticas Vitais - Mortalidade e Nascidos Vivos). Com os avanços no controle das doenças infecciosas (Informações Epidemiológicas e Morbidade) e com a melhor compreensão do conceito de saúde e de seus determinantes populacionais, a análise da situação sanitária passou a incorporar outras dimensões do estado de saúde, como por exemplo: dados de morbidade, incapacidade, acesso a serviços, qualidade da atenção, condições de vida e fatores ambientais. Estas dimensões passaram a ser métricas utilizadas na construção de Indicadores de Saúde, que se traduzem em informação relevante para a quantificação e a avaliação da saúde brasileira (DATASUS A, 2015).

## Ferramentas de *Business Intelligence* do DATASUS

O DATASUS armazena seus dados em banco de dados utilizando conceitos de modelagem dimensional. Eles são desenvolvidos e carregados com séries históricas, provenientes de sistemas transacionais estruturados ou não, facilitando o desenvolvimento de projetos de *Business Intelligence*. A Tabela 2 mostra alguns projetos de BI ordenados por situação, utilizando os dados disponíveis no DATASUS. Os demais projetos podem ser visualizados no Anexo A. A situação indica projetos que já foram concluídos e outros que estão em fase de desenvolvimento (DATASUS B, 2015).

Tabela - Projetos de *Business Intelligence* desenvolvidos através do DATASUS

| Projeto | Descrição | Situação |
| --- | --- | --- |
| SISCAN | Gestão e Acompanhamento do Programa de Câncer de Mama e de Colo de Útero. | Construção |
| MAIS MÉDICOS | Gestão dos Participantes do Programa Mais Médicos. | Produção |
| FARMÁCIA POPULAR | Gestão e acompanhamento da distribuição de medicamentos do Programa Farmácia Popular. | Produção |
| DENGUE | Gestão e acompanhamento da evolução dos casos de DENGUE no Brasil. | Produção |
| BOLSA ALIMENTAÇÃO | Acompanhamento do Sistema Bolsa Alimentação. | Produção |

Fonte: DATASUS B, 2015.

Observando o Anexo A, é possível analisar que existe uma grande variedade de ferramentas de *Business Intelligence* para atender as mais diversas necessidades de obter informações na área da Saúde. Porém o acesso a estas ferramentas fica restrito a usuários credenciados e profissionais da Saúde, sendo necessária a autenticação através de usuário e senha, os quais são fornecidos pelo gestor de cada ferramenta em específico.

Diante desde cenário, observando a Tabela 2, é possível identificar que o DATASUS dispõe de diversos sistemas de *Business Intelligence* que atendem as mais diversas áreas da saúde. No entanto ainda existe uma grande variedade de dados que não foram explorados por estas áreas. Através do aplicativo TABNET, disponível em seu portal, o DATASUS disponibiliza diversas tabelas bidimensionais. De posse destas tabelas, é possível extrair seus dados e importá-los para sistemas de *Business Intelligence*. Desta forma, consultas, cruzamentos e análises multidimensionais podem ser realizadas, fornecendo excelentes resultados aos profissionais da área de saúde e também à população brasileira.

## Definição dos dados para análises

Com base nas pesquisas e análises realizadas nos dados disponíveis no portal do DATASUS, foram utilizados os dados extraídos da Pesquisa Nacional de Saúde de 2013 (PNS) para a realização deste trabalho. A PNS é uma pesquisa domiciliar, de abrangência nacional, com uma amostra de 80.000 domicílios em 1.600 municípios brasileiros. Sua realização é feita em parceria com o Instituto Brasileiro de Geografia e Estatística (IBGE) e acontece a cada 5 anos.

A composição dessa pesquisa é formada por três questionários: o domiciliar, referente às características do domicílio; o relativo a todos os moradores do domicílio; e o individual, a ser respondido por um morador de 18 anos ou mais do domicílio. Seus dados foram agrupados em módulos, conforme é exibido na Tabela 3 (DATASUS D, 2015).

Tabela - Módulos da Pesquisa Nacional da Saúde 2013

|  |  |
| --- | --- |
| Módulo | Descrição |
| A | Informações do Domicílio. |
| B | Visitas domiciliares de Equipe de Saúde da Família e Agentes de Endemias. |
| G | Pessoas com deficiências. |
| I | Cobertura de Plano de Saúde. |
| J | Utilização de Serviços de Saúde. |
| K | Saúde dos indivíduos com 60 anos ou mais. |
| L | Crianças com Menos de 2 Anos. |
| N | Percepção do estado de saúde. |
| O | Acidentes e violências - Trânsito e trabalho. |
| O | Acidentes e violências - Violências e agressões. |
| P | Estilo de vida - Uso de álcool e tabagismo. |
| P | Estilo de vida - Consumo de alimentos. |
| P | Estilo de vida - Atividade física e hábito de assistir televisão. |
| Q | Doenças crônicas - Hipertensão arterial. |
| Q | Doenças crônicas – Diabetes. |
| Q | Doenças crônicas - Colesterol, Cardiovasculares (DCV), AVC e Asma. |
| Q | Doenças crônicas - Artrite/reumatismo, coluna, DORT, depressão e outras mentais. |
| Q | Doenças crônicas - Pulmão, câncer, insuficiência renal e outras. |
| K | Cobertura de mamografia entre mulheres de 50 anos ou mais. |
| R | Saúde da Mulher. |
| S | Atendimento Pré-natal. |
| U | Saúde Bucal. |
| X | Atendimento médico. |

Fonte: DATASUS D, 2015.

Porém, devido à grande quantidade de módulos dentro da PNS 2013, foi necessário escolher apenas uma fatia para a elaboração deste trabalho. Para esta escolha foram utilizados dois critérios: ter informações de interesse social da grande maioria da população nacional; e não depender de um especialista do domínio da área da saúde para compreender os dados contidos.

As mortes por causas externas são todos os tipos de acidentes e violências. Elas são consideradas um problema de saúde pública na atualidade, principalmente em países em desenvolvimento (MASCARENHAS; BARROS, 2015). Dentro das mortes por causas externas estão os acidentes de trânsito, os homicídios e os suicídios. Estas três causas em conjunto respondem por cerca de dois terços das mortes por causas externas no Brasil. Ainda dentro deste conjunto, as taxas são consideravelmente mais elevadas na população de adultos e jovens, principalmente do sexo masculino (DATASUS C, 2015).

Segundo a *World Health Organization* (WHO) - Organização Mundial da Saúde (OMS), 5,1% da carga global de doenças estão atribuídas ao álcool. Além disso, a cada ano, ocorrem 3,3 milhões de mortes em decorrência do consumo nocivo do álcool. Também há evidências científicas de que uma proporção importante da carga de doença atribuível ao álcool é decorrente de lesões não intencionais e intencionais, incluindo-se aquelas devidas a acidentes de trânsito e violências (WHO, 2014).

Dadas as considerações acima, dentre todos os módulos da PNS 2013 disponíveis no portal do DATASUS, foram extraídos os dados referentes aos módulos: “Acidentes e violências - Trânsito e trabalho - Módulo O”; e “Estilo de vida - Uso de álcool e tabagismo - Módulo P”.

A razão pela qual foi escolhida a última Pesquisa Nacional da Saúde (PNS) como fonte de dados para o desenvolvimento do ambiente de BI, proposto neste trabalho, dentre todos os dados acessíveis e disponibilizados no DATASUS, foi pela forma como a mesma é apresentada em seu portal. Através dele não é possível realizar cruzamentos entre os seus módulos e também suas consultas são exibidas em formato de tabelas bidimensionais, mostrando as informações em linhas e colunas.

Também foi levada em consideração a organização dos dados da PNS 2013, que estão distribuídos e classificados em diversos módulos. O que faz com que eles tenham um grande potencial para serem analisados através do apoio de um conjunto de ferramentas de BI. Com elas é possível exibi-los e realizar diversos tipos de cruzamentos entre os módulos e obter análises mais detalhadas e dinâmicas sobre os dados.

# DESENVOLVIMENTO DO *DATA WAREHOUSE* E DOS PROCESSOS ETL

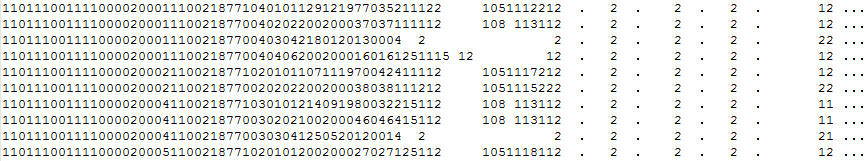
Após a definição dos dados, o próximo passo para a implementação de um ambiente de BI foi o desenvolvimento dos processos de ETL e modelagem do *Data Warehouse*. Para essa etapa, que será explanada no decorrer deste capítulo, foram utilizados os conceitos, técnicas e processos mostrados na seção 1.1 do primeiro capítulo deste trabalho.

Para realizar a armazenagem e criar o *Data Warehouse* dos dados extraídos dos processos de ETL, optou-se pelo uso do Sistema de Gerenciamento de Banco de Dados (SGBD) MySQL. Um banco de dados relacional que utiliza a linguagem SQL (*Structured Query Language*). Tal escolha justifica-se pelo fato dele ser *open source* e não possuir nenhuma restrição de uso.

## Etapas da criação do *Data Warehouse* e processos ETL

Para o desenvolvimento do *Data Warehouse*, foram realizados os processos e técnicas de ETL, utilizando os dados obtidos no site do IBGE em formato de arquivos microdados. Este formato consiste no menor nível de desagregação dos dados de uma pesquisa, retratando, sob a forma de códigos numéricos, o conteúdo dos questionários. Desta forma, preserva-se o sigilo estatístico com vistas à não individualização das informações, conforme é possível observar um trecho na Figura 11. A partir deles usuários especializados, com conhecimento em programação, poderão realizar a leitura dos dados, o cruzamento em diferentes agregações geográficas, e a elaboração de múltiplas tabulações segundo sua perspectiva pessoal de interesse.

Figura - Arquivo de microdados “PESPNS2013.txt”

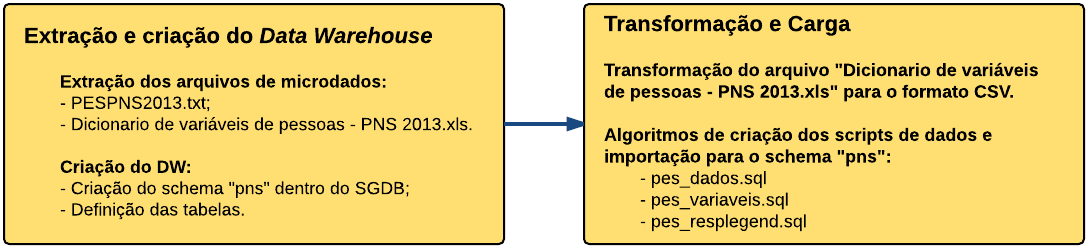


Fonte: Autor, 2016.

Os arquivos de microdados são acompanhados de uma documentação auxiliar, conforme é possível observar na Figura 14. Sua função é fornecer os nomes, os códigos e as descrições das variáveis e suas categorias, complementada, quando necessário, dos elementos necessários para o cálculo dos erros amostrais (IBGE, 2016).

A fim de ajudar na compreensão dos processos executados nesta seção do trabalho, foi elaborado um fluxograma, conforme demonstra a Figura 12. Nele foram resumidas as etapas relatadas nas próximas duas subseções.

Figura – Resumo das etapas: Extração e criação do *Data Warehouse*; Transformação e Carga

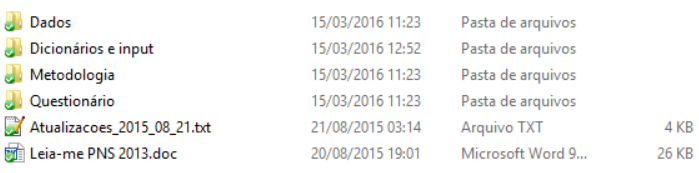


Fonte: Autor, 2016.

### Extração e criação do *Data Warehouse*

Durante esta etapa do processo de ETL, foi realizado o *download* do arquivo microdados da PNS 2013, conforme demonstra a Figura 13 com seu conteúdo descompactado. Além disso, foi realizada a modelagem do banco de dados, para que fosse possível armazenar os dados e informações auxiliares extraídos do arquivo de microdados.

Figura - Arquivo microdados.zip descompactado

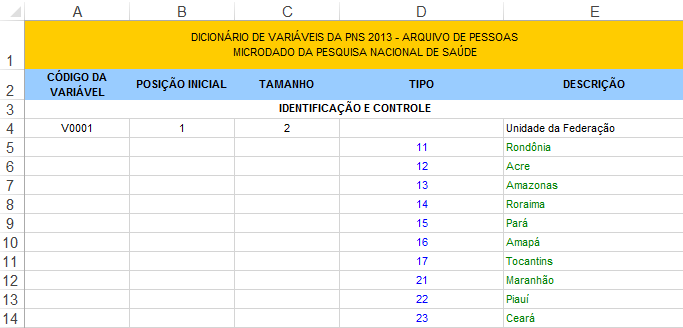


Fonte: Autor, 2016.

Para a extração dos dados necessários para realizar a proposta deste trabalho, foram identificados dois arquivos: ”PESPNS2013.txt”, disponível no diretório “Dados”; e “Dicionário de variáveis de pessoas - PNS 2013.xls”, disponível no diretório “Dicionários e input”.

* Arquivo “PESPNS2013.txt”: responsável por armazenar as respostas de cada entrevistado durante a pesquisa em formato microdados. Observando a Figura 11, é possível visualizar um trecho do seu conteúdo.
* Arquivo “Dicionário de variáveis de pessoas - PNS 2013.xls”: planilha responsável por armazenar os nomes, as posições e o tamanho das variáveis de cada resposta contida em cada registro do arquivo de microdados. Seu conteúdo é apresentado conforme trecho demonstrado pela Figura 14.

Figura – Trecho do Dicionário de variáveis do arquivo de microdados da PNS 2013



Fonte: Autor, 2016.

De posse destes dois arquivos foi possível iniciar a criação e modelagem do banco de dados. Para isso, através do script “0-schema e usuario.sql”, disponível no Apêndice A, foi criado um *schema*, identificado como “pns”, responsável por armazenar as tabelas e seus respectivos dados. As tabelas criadas foram as seguintes:

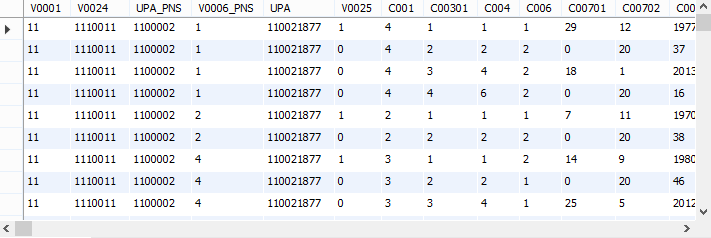
* pes\_dados: responsável por armazenar a resposta de todas as perguntas do questionário da PNS. A cada registro são armazenadas todas as respostas de cada entrevistado;
* pes\_variaveis: responsável por armazenar as perguntas contidas no questionário da PNS. Cada registro armazenado no atributo “código”, corresponde a um nome de atributo contigo na tabela “pes\_dados”;
* pes\_resplegend\_”n”: responsável por armazenar a legenda de cada resposta contida nos atributos da tabela dados, onde “n”, corresponde ao nome do atributo da tabela dados.

### Transformação e Carga

Após a realização da etapa de extração e modelagem, onde foram identificados os arquivos: ”PESPNS2013.txt” e “Dicionário de variáveis de pessoas - PNS 2013.xls”. O próximo passo, utilizando a linguagem de programação C Sharp, foi o desenvolvimento de um algoritmo capaz de realizar a leitura e a interpretação destes dois arquivos. Através dele foi possível gerar os *scripts* SQL para a importação no banco de dados, conforme é demonstrado:

* **Transformação do arquivo “Dicionário de variáveis de pessoas - PNS 2013.xls"**: consiste apenas em abrir o arquivo no Microsoft Excel e salvá-lo em um arquivo separado por vírgulas (Dicionário de variáveis de pessoas - PNS 2013.csv).
* **Carga do arquivo “Dicionário de variáveis de pessoas - PNS 2013.csv"**: nesta etapa o algoritmo carrega todo o conteúdo do arquivo em uma tabela temporária identificada como “dtDicVarPes”, percorrendo todas as suas linhas em busca de padrões definidos para identificar: o nome da variável; sua descrição; o tamanho e a posição inicial dentro do arquivo de microdados “PESPNS2013.txt”, o qual é ilustrado pela Figura 11.
* ***Script* “pes\_dados.sql”**: consiste em duas etapas: criar a tabela pes\_dados e gerar seus *inserts*. Para a primeira etapa da geração do *script*, o algoritmo insere em sua primeira linha o comando responsável pela criação da tabela “pes\_dados”: “CREATE TABLE pes\_dados (”. Para a geração das linhas subsequentes do *script* e também a continuação do comando para criar a tabela, ainda dentro da primeira etapa, foi necessário percorrer e identificar todas as variáveis armazenadas dentro da tabela temporária “dtDicVarPes”. Para cada variável encontrada é adicionada uma nova linha no *script* com o conteúdo “nome\_da\_variavel int(20)”. Após a conclusão da geração do comando para criar a tabela foi realizada a segunda etapa, que consiste em realizar a carga do arquivo de microdados “PESPNS2013.txt” e percorrer todas as suas linhas. Para cada linha encontrada no arquivo, o algoritmo percorre toda a tabela temporária “dtDicVarPes”, para cada variável encontrada é identificado seu nome, sua posição inicial e seu tamanho e realizada a leitura da linha atual. De posse dessas informações é possível gerar o comando de inserção de dados na tabela “pes\_dados”. Para cada linha do arquivo “PESPNS2013.txt” é gerado um comando, consequentemente uma nova linha no *script*: “INSERT INTO pes\_dados (nome\_dos\_campos) VALUES (valores\_dos\_campos)”. A Figura 15 ilustra o resultado de um comando de seleção de dados (SELECT \* FROM pes\_dados) na tabela “pes\_dados”, após a execução do *script*.

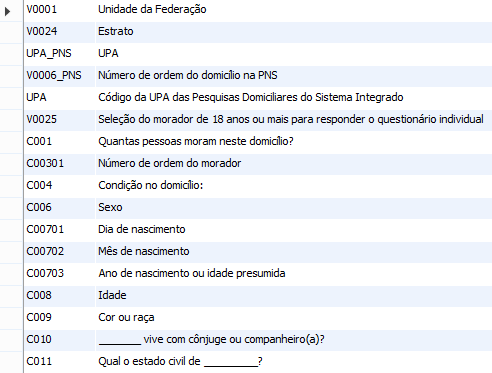
Figura - Comando *Select* na tabela "pes\_dados"

****

Fonte: Autor, 2016.

* ***Script* “pes\_variaveis.sql”**: nesta etapa o algoritmo insere a primeira linha no *script*, que é o comando SQL responsável por criar a tabela “pes\_variaveis” “CREATE TABLE pes\_variaveis (codigo varchar(20), descricao VARCHAR(400) NULL))”. Após, é realizada a leitura da tabela temporária “dtDicVarPes” e, para cada variável encontrada, é gerado um novo comando SQL, responsável por inserir os dados na tabela: “INSERT INTO pes\_variaveis (codigo, descricao) VALUES (nome\_da\_variavel, descricao\_da\_variavel)”. A Figura 16 demonstra o resultado do comando *Select* na tabela “pes\_variaveis”, após a execução do *script* criado.

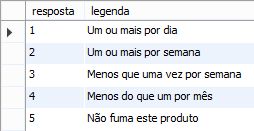
Figura - Comando *Select* na tabela "pes\_variaveis"

****

Fonte: Autor, 2016.

* ***Script* “pes\_resplegend.sql”**: na última fase do algoritmo de transformação, são identificados os nomes das variáveis e suas respostas. Para cada variável encontrada é criada uma tabela através do comando SQL “CREATE TABLE pes\_resplegend\_nome\_da\_variavel (resposta varchar(20), legenda VARCHAR(400) NULL)” e seus *inserts*. Após a execução do *script*, a Figura 17 apresenta o resultado do comando de seleção na tabela “pes\_resplegend\_p05404”, representando uma das “n” ou “760” tabelas geradas nesta fase.

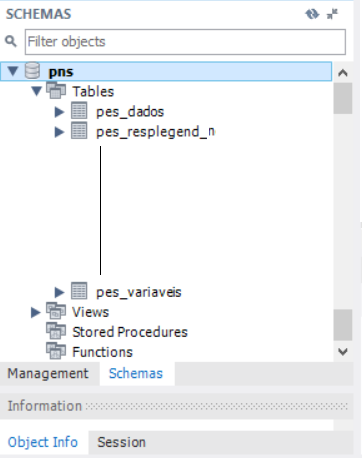
Figura - Comando *Select* na tabela “pes\_resplegend\_p05404”



Fonte: Autor, 2016.

A Figura 18 ilustra o esquema “pns” após a geração e execução de todos os *scripts* dentro do banco de dados (*schema*). Reforçando que a representação de todas as 760 tabelas, responsável por armazenar a legenda de cada resposta contida nos atributos da tabela “pes\_dados”, está denotada pela tabela “pes\_resplegend\_n”.

Figura - *Schema* “pns”



Fonte: Autor, 2016.

Durante a análise e avaliação dos dados dentro do DW, foi detectada a necessidade de realizar mais algumas alterações na estrutura de dados e tabelas. Foram elas:

* Criação de uma tabela auxiliar “aux\_faixa\_etaria”: responsável por armazenar a faixa etária dos entrevistados de acordo com a sua idade. A definição para cada grupo está de acordo com aquela já utilizada dentro do TABNET pelo DATASUS: de 0 a 17 anos, de 18 a 24 anos, de 25 a 39 anos, de 40 a 59 anos e de 60 anos ou mais. Para a criação desta tabela foi desenvolvido o *script* “4-aux\_faixa\_etaria.sql”, que também pode ser visualizado no Apêndice A.
* Atualização da tabela “pes\_resplegend\_V0001”: responsável por armazenar a Unidade da Federação do entrevistado. Nessa tabela foi adicionado um novo atributo denominado “regiao”, sua função é armazenar o nome da região a qual pertence cada Unidade da Federação, de acordo com o IBGE. Para realizar está alteração foi criado um script “5-pes\_resplegend\_v0001.sql”, que também pode ser visualizado no Apêndice A.

Após a criação, inserções e ajustes de todas as tabelas, foi contabilizado um total de 763 tabelas. Seus registros ocupam um espaço total de 138,36 *Megabytes* (MB) dentro do banco de dados, destacando que a tabela “pes\_dados”, é responsável por 89% do espaço total utilizado.

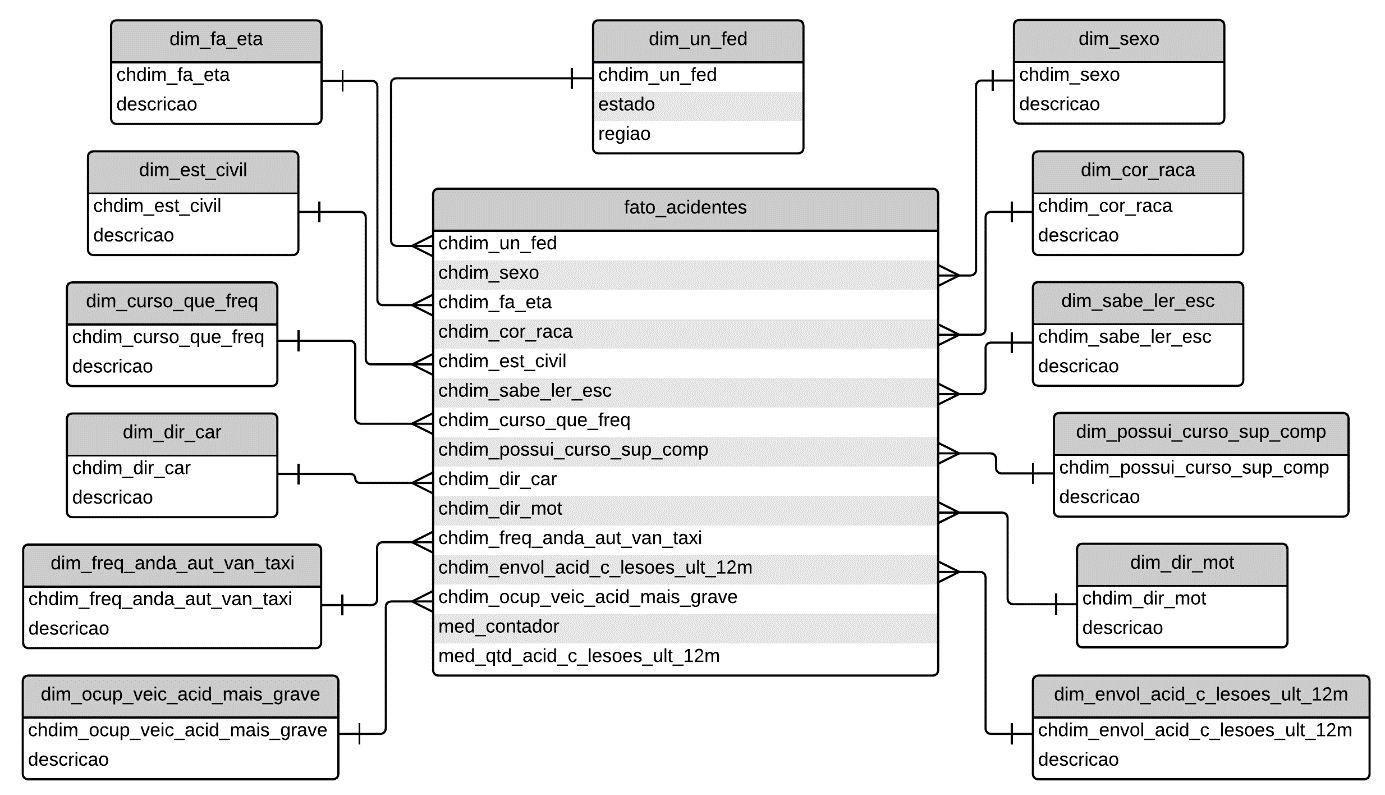
## Modelagem multidimensional do *Data Warehouse*

Nesta seção do trabalho, utilizando os dados armazenados dentro do DW, foram desenvolvidas algumas *views* para apresentar os dados armazenados como um cubo. Para isso foram criadas duas *views* fato, uma para representar o cubo acidentes e outra para o cubo uso\_alcool. Além disso, para cada uma das duas, foram criadas também outras *views* para representar suas dimensões.

### *View* “fato\_acidentes”

Para representar o cubo “acidentes”, utilizando dados do módulo "Acidentes e violências - Trânsito e trabalho" da PNS 2013, foi desenvolvida uma *view* para representar o fato e mais treze *views* para representar suas dimensões. Neste modelo foram inseridos apenas os atributos para os dados relacionados a acidentes de trânsito. A Figura 19, mostra as ligações das *views* dimensões com a *view* fato no centro do modelo estrela.

Figura – Modelo estrela da *view* “fato\_acidentes” e suas dimensões



Fonte: Autor, 2016.

Os atributos da *view* fato\_acidentes podem ser melhores compreendidos observando a Tabela 4. Nela é possível identificar quais são os atributos que armazenam as chaves para as dimensões e quais os que armazenam as medidas. Para as medidas, que são usadas para representar as métricas de interesse, seus nomes iniciam-se pelo termo “med\_”, abreviando “medida”. Com relação as chaves de dimensões, seus nomes iniciam-se com a expressão “chdim\_”, abreviatura para “chave dimensão”.

Tabela - Fato “acidentes”: detalhamento dos atributos

| Nome | Descrição |
| --- | --- |
| chdim\_un\_fed | Unidade da Federação e Região |
| chdim\_sexo | Sexo |
| chdim\_fa\_eta | Faixa etária |
| chdim\_cor\_raca | Cor ou raça |
| chdim\_est\_civil | Estado civil |
| chdim\_sabe\_ler\_esc | Sabe ler e escrever |
| chdim\_curso\_que\_freq | Curso que frequenta |
| chdim\_possui\_curso\_sup\_comp | Possui algum curso superior de graduação completo |
| chdim\_dir\_car | Dirige carro. |
| chdim\_dir\_mot | Dirige motocicleta. |
| chdim\_freq\_anda\_aut\_van\_taxi | Frequência de uso de automóvel, van ou táxi. |
| chdim\_envol\_acid\_c\_lesoes\_ult\_12m | Se envolveu em acidente de trânsito com lesões corporais nos últimos 12 meses. |
| chdim\_ocup\_veic\_acid\_mais\_grave | Ocupação no acidente de trânsito mais grave ocorrido nos últimos 12 meses. |
| med\_contador | Contador que é incrementado de acordo com as combinações das dimensões. |
| med\_qtd\_acid\_c\_lesoes\_ult\_12m | Quantidade de acidente de trânsito com lesões corporais nos últimos 12 meses. |

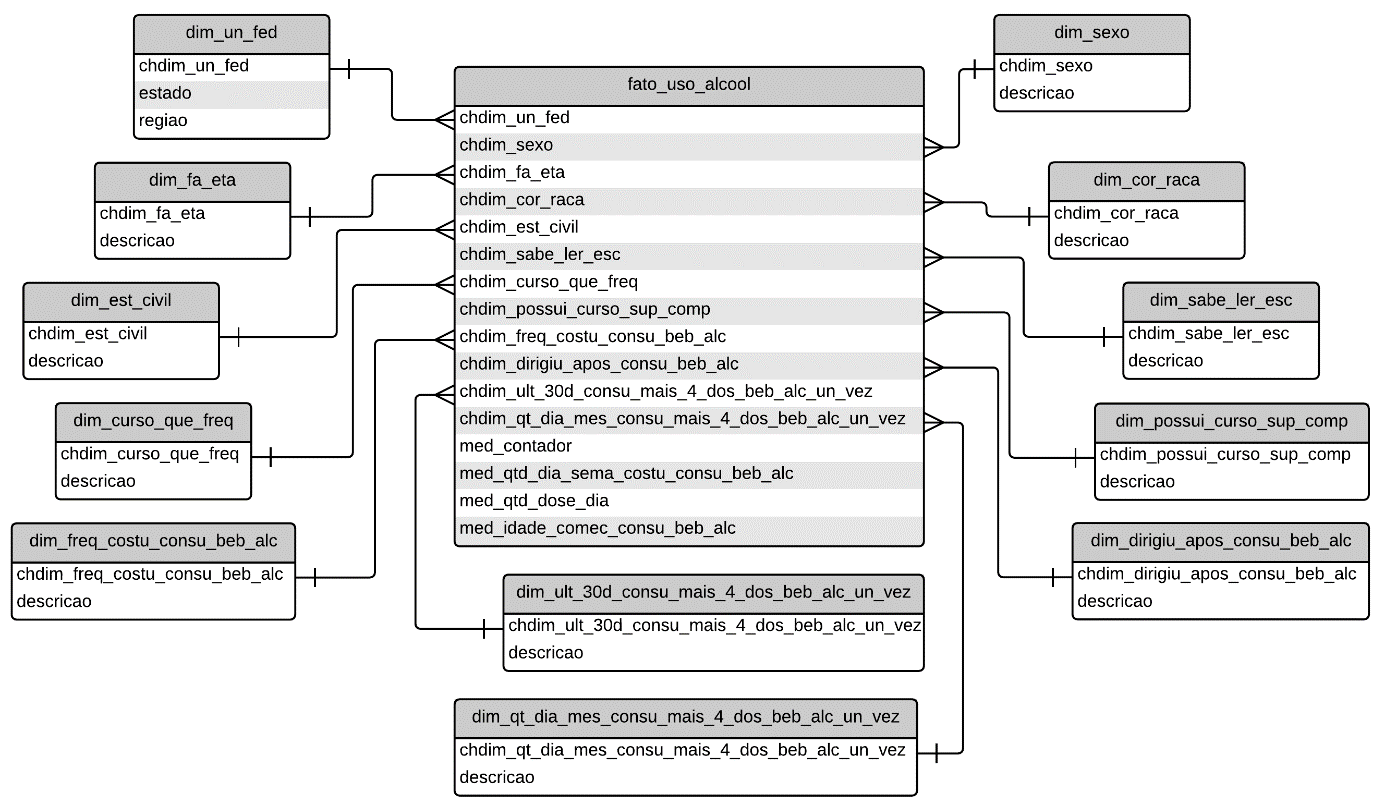
Fonte: Autor, 2016.

Através deste conjunto de *views*, modeladas de forma multidimensional, foi possível representar os dados do módulo "Acidentes e violências - Trânsito e trabalho" em formato de um cubo. Desta forma, é possível realizar análises mais detalhadas, como por exemplo: identificar qual a faixa etária dos entrevistados se envolveu em mais acidentes de trânsito. Permitindo também que sejam facilmente adicionados mais detalhes à consulta, como por exemplo, cruzar os dados: sexo; estado civil; nível de instrução. Para mais detalhes sobre os registros de cada uma das *views* fato e dimensões, consultar o Apêndice B.

### View “fato\_uso\_alcool”

Para apresentar o cubo “uso\_alcool”, mostrando os dados do módulo "Estilo de vida - Uso de álcool e tabagismo" da PNS 2013, foi criada uma *view* para representar o fato “uso\_alcool” e mais doze *views* para as suas dimensões. Neste modelo foram inseridos apenas os atributos para os dados relacionados ao uso de álcool, conforme é possível observar na Figura 20. Nela é exibida a *view* fato no centro do modelo estrela com suas ligações às *views* dimensões.

Figura – Modelo estrela da *view* “fato\_uso\_alcool” e suas dimensões



Fonte: Autor, 2016.

Conforme é observado na Figura 20, a *view* “fato\_uso\_alcool” contém em seus atributos as chaves para as *views* dimensões e também as suas medidas. Seus atributos podem ser melhor compreendidos observando-se a Tabela 5.

Tabela - Fato “uso\_alcool”: detalhamento dos atributos

| Nome | Descrição |
| --- | --- |
| chdim\_un\_fed | Unidade da Federação e Região. |
| chdim\_sexo | Sexo. |
| chdim\_fa\_eta | Faixa etária. |
| chdim\_cor\_raca | Cor ou raça. |
| chdim\_est\_civil | Estado civil. |
| chdim\_sabe\_ler\_esc | Sabe ler e escrever. |
| chdim\_curso\_que\_freq | Curso que frequenta. |
| chdim\_possui\_curso\_sup\_comp | Possui algum curso superior de graduação completo. |
| chdim\_freq\_costu\_consu\_beb\_alc | Frequência que costuma consumir alguma bebida alcoólica. |
| chdim\_dirigiu\_apos\_consu\_beb\_alc | Dirigiu após consumir álcool. |
| chdim\_ult\_30d\_consu\_mais\_4\_dos\_beb\_alc\_un\_vez | Nos últimos 30 dias, consumiu 4 ou mais doses de bebida alcoólica em uma única ocasião. |
| chdim\_qt\_dia\_mes\_consu\_mais\_4\_dos\_beb\_alc\_un\_vez | Nos últimos 30 dias, quantidade de dias que consumiu 4 ou mais doses de bebida alcoólica em uma única ocasião. |
| med\_contador | Contador que é incrementado de acordo com as combinações das dimensões. |
| med\_qtd\_dia\_sema\_costu\_consu\_beb\_alc | Quantidade de dias por semana que costuma consumir alguma bebida alcoólica. (0 = Nunca ou menos de uma vez por semana). |
| med\_qtd\_dos\_dia | Quantidade de doses por dia. (Para os dias que consome bebida alcoólica) |
| med\_idade\_comec\_consu\_beb\_alc | Idade que começou a consumir álcool |

Fonte: Autor, 2016.

Através das *views* mostradas na na Figura 20, foi possível representar os dados do módulo "Estilo de vida - Uso de álcool e tabagismo" como um cubo. Permitindo realizar análises mais detalhadas, como por exemplo: identificar qual a faixa etária dos entrevistados já dirigiu após consumir álcool. Possibilitando também que sejam adicionados mais detalhes a consulta, como por exemplo, cruzar os dados: sabe ler e escrever; estado civil; se envolveu em acidentes de trânsito ou não nos últimos doze meses. Para mais detalhes sobre os registros de cada uma das *views* fato e dimensões, consultar o Apêndice B.

# IMPLEMENTAÇÃO DA FERRAMENTA DE BI

Para a implementação da ferramenta de BI, foi utilizado o ambiente definido na seção 2.5 deste trabalho. Primeiramente foi realizada a preparação e instalação do ambiente *Pentaho Community*, em seguida foi configurada a conexão do ambiente com o *Data Warehouse* e a criação e configuração dos cubos de dados “acidentes”, “uso\_alcool” e “acid\_vs\_uso\_alc”. Utilizando *plug-ins* OLAP, foram desenvolvidas as visualizações dos cubos e, por fim, foram extraídas percepções de três usuários que manipularam esses cubos.

## Preparação e instalação do ambiente

Nesta etapa da implementação é abordado como foi o processo de instalação do ambiente do *Pentaho Community*. Necessário para criar e desenvolver os cubos e visões dos dados extraídos, além de permitir a realização de cruzamentos e consultas OLAP. Para isso foram definidas oito etapas, as quais serão explanadas no decorrer desta seção, conforme é possível observar na Figura 21.

Figura - Etapas da preparação e instalação do *Pentaho Community*



Fonte: Autor, 2016.

Para as três primeiras etapas foi realizado um levantamento dos pré-requisitos necessários. Além de fazer o download dos arquivos de instalação de todos os itens necessários para a execução e instalação do ambiente, conforme é detalhado:

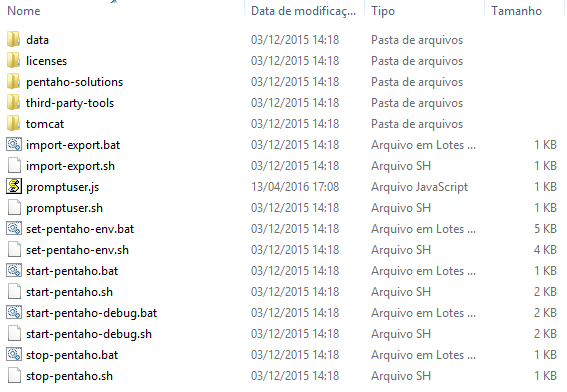
* Sistema Operacional Windows, Linux ou Mac.
* Kit de desenvolvimento Java “Sun Java Developer Kit (JDK)”, disponível em: http://www.oracle.com/technetwork/java/javase/downloads/.
* Arquivo de instalação do *Pentaho Community*, que pode ser encontrado em: http://community.pentaho.com/.

Após o download dos arquivos necessários, optou-se por utilizar o sistema Operacional Windows. Tal escolha deve-se ao fato do autor não possuir conhecimento dos demais Sistemas Operacionais e de que o propósito desta etapa é de apenas implementar o ambiente, independente do Sistema Operacional.

Para a quarta e quinta fase desta seção, seguindo a Figura 21, foi realizada a instalação do Kit de Desenvolvimento Java dentro do Sistema Operacional. Para isso foi executado o arquivo de instalação e seguido os passos do próprio instalador. Além disso, para que o Java pudesse executar o *Pentaho Community*, foi necessário realizar algumas configurações dentro das variáveis de ambiente do Sistema Operacional. Para isso foi adicionada uma nova variável de usuário “JAVA\_HOME”, a qual foi apontada para o diretório de instalação do JAVA e para a variável de sistema existente “PATH”, seu valor foi complementado, adicionando um apontamento para o diretório de binários do JAVA.

Dando sequência nos processos definidos na Figura 21, após a preparação do Sistema Operacional, foi realizada a instalação da ferramenta de BI. Para isso, foi realizada a extração do arquivo compactado, obtido no site do Pentaho para dentro do diretório criado “C:\Pentaho\”, conforme ilustra Figura 22.

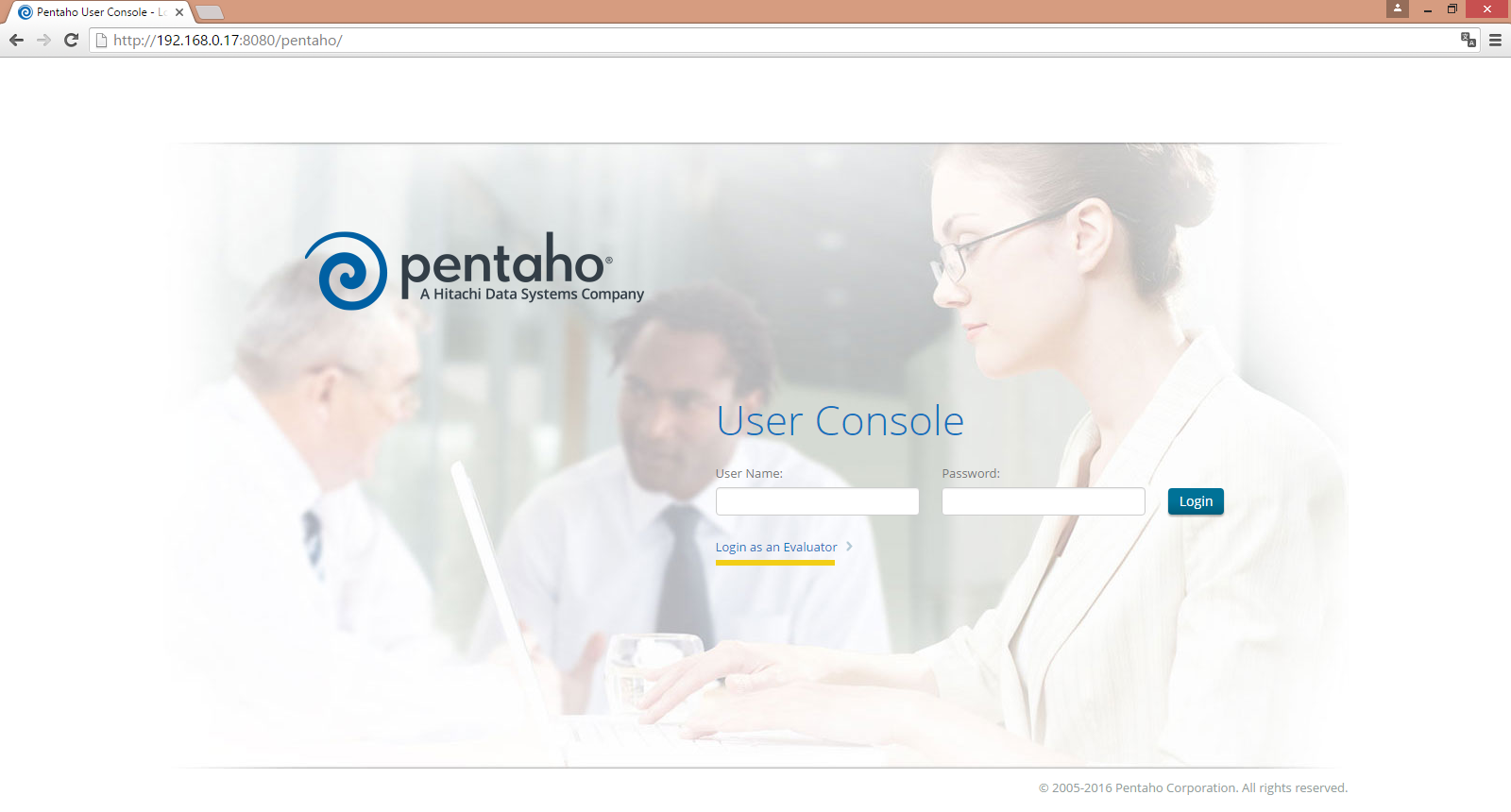
Figura - Diretório do *Pentaho Community*



Fonte: Autor, 2016.

O *Pentaho Community* não precisou de procedimentos avançados de instalação, sendo necessário apenas executar o arquivo “start-pentaho.bat” para inicializar a plataforma e “stop-pentaho.bat” para finalizá-la, ambos disponível no diretório do *Pentaho Community*. Seu acesso é realizado através de um navegador de internet na porta 8080, utilizando o endereço *Internet Protocol* (IP) do computador, que está executando a plataforma. Conforme é demonstrado Figura 23, utilizando o navegador Google Chrome. Para o primeiro acesso à ferramenta, é necessário utilizar seu usuário e senha padrão, conforme consta em sua documentação disponível no seu endereço eletrônico: usuário*: admin;* senha*: password*.

Figura - Desenvolvimento: Tela inicial do *Pentaho Community*



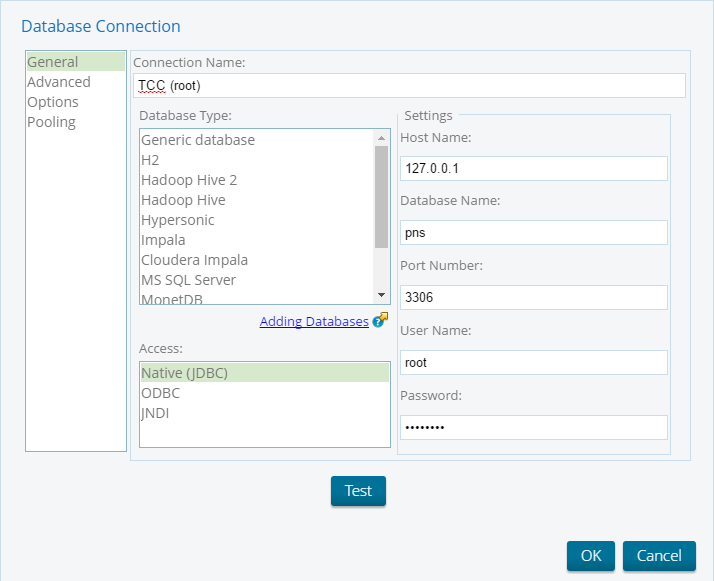
Fonte: Autor, 2016.

## Conexão da ferramenta com o *Data Warehouse*

Durante a etapa de implementação da criação da conexão do *Pentaho Community* com o DW, foi necessário definir um nome para a conexão e também informar as credenciais de acesso, conforme ilustra a Figura 24. São elas:

* *Connection Name*: DWPNS (definido pelo autor);
* *Database Type*: MySQL (SGBD utilizado);
* *Access*: *Native* (JDBC);
* *Host Name*: localhost;
* *Data Base* Name: pns;
* *Port Number*: 3306;
* *User Name*: root;
* *Password*: feevale.

Figura - *Pentaho Community*: conexão com o banco de dados com usuário root

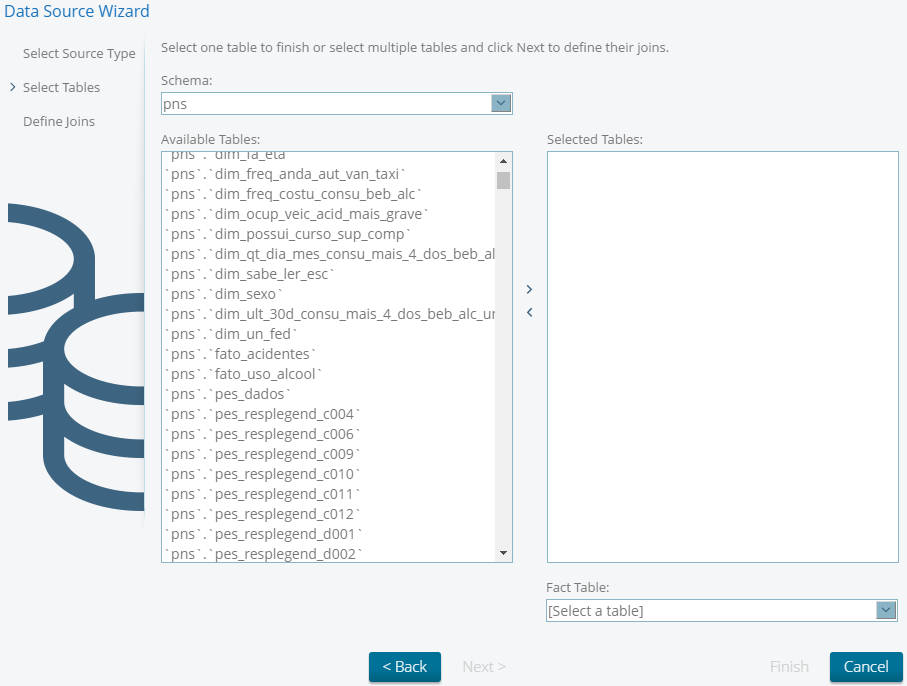


Fonte: Autor, 2016.

## Criação e configuração dos cubos (*Data Sources)*

Durante o desenvolvimento desta etapa, foi identificada a necessidade de criar um usuário com acesso limitado somente às *views* criadas na seção 4.2 do capítulo 4. Tal razão deve-se ao fato de que a conexão criada com o usuário “root” possui privilégios administrativos, fazendo com que ela tenha acesso a todas as 763 tabelas do *schema* “pns”, dificultando a localização e seleção das *views* para o desenvolvimento dos cubos, conforme é ilustrado pela Figura 25.

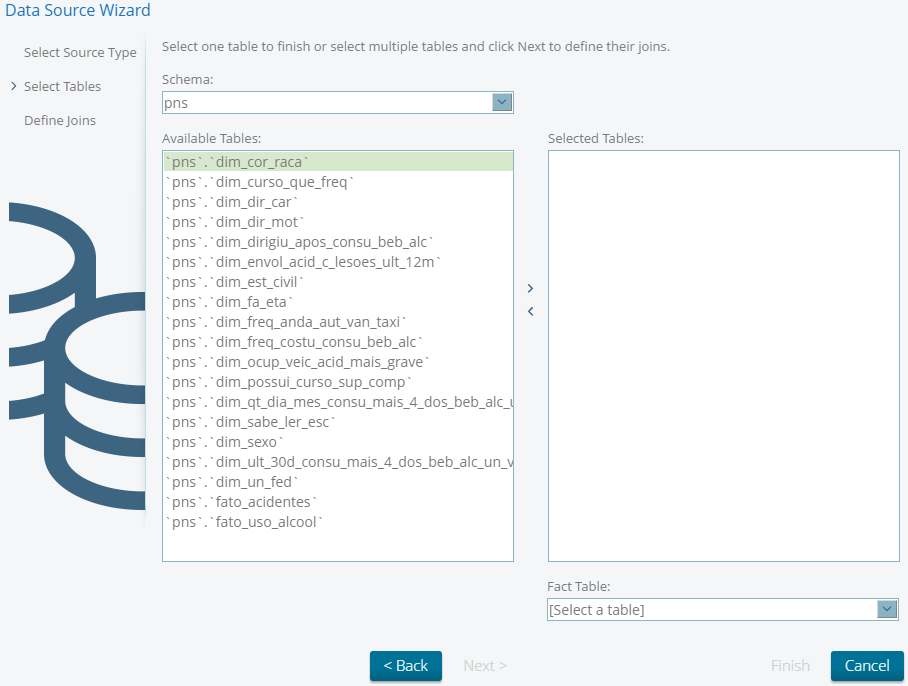
Figura - Seleção das tabelas fato e dimensões com usuário root



Fonte: Autor, 2016.

Para melhorar esta visualização e seleção das *views* durante a criação dos cubos, foi criado um novo usuário no banco de dados com permissão de acesso somente as *views* necessárias para o desenvolvimento dos cubos. Com este novo usuário foi criada uma nova conexão ao *Data Warehouse*, tal como foi feito na seção 5.2. Através deste novo usuário foi possível realizar a seleção das *views* em um ambiente menos poluído, pois foram listadas somente as *views* necessárias para o desenvolvimento dos cubos propostos, conforme pode-se observar na Figura 26.

Figura - Seleção das tabelas fato e dimensões com o usuário limitado

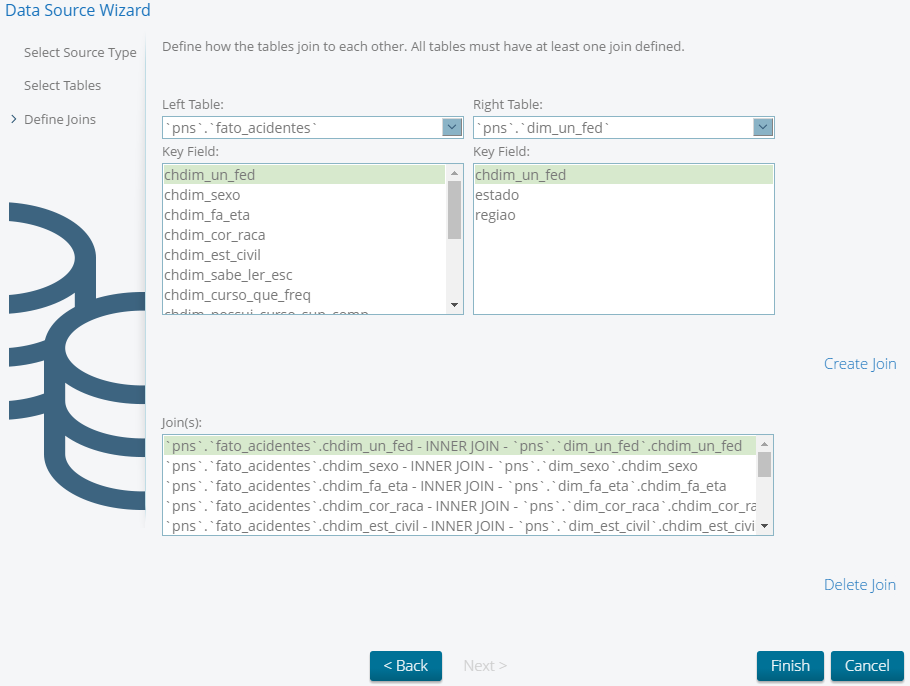


Fonte: Autor, 2016

### Cubo “acidentes”

O primeiro passo para a criação do cubo “acidentes”, dentro do ambiente do *Pentaho Community*, foi a seleção das *views* fato e dimensões, conforme é demonstrado pela Figura 26. Em seguida foram realizadas as junções de cada *view* dimensão com a *view* fato, conforme é ilustrado na Figura 27. Uma operação de junção pode ser exemplificada observando-se a associação do atributo “chdim\_und\_fed” da tabela fato “fato\_acidentes” com o atributo “chdim\_un\_fed” da tabela dimensão “dim\_un\_fed”.

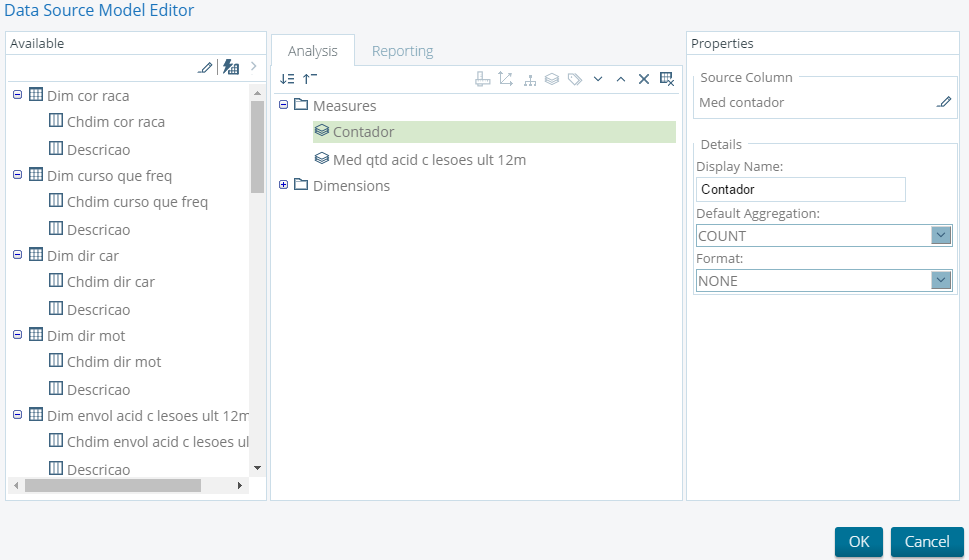
Figura – Criação dos *joins* do cubo “acidentes”



Fonte: Autor, 2016.

Após a finalização do processo de criação do cubo, onde foram selecionadas todas as *views* necessárias para representá-lo, foi preciso realizar a configuração dos atributos de cada uma das *views*. Para cada medida selecionada foi preciso definir o nome de exibição e informar o tipo de agregação para cada valor armazenado na medida, conforme é demonstrado na Figura 28. Com relação as dimensões, para cada uma delas foi preciso adicionar apenas os seus atributos que ficarão visíveis, além de definir um nome amigável e sugestivo para que seja possível realizar uma melhor interpretação no momento de visualização do cubo.

Figura – *Pentaho Community*: seleção e configuração das medidas do cubo



Fonte: Autor, 2016.

Para as medidas que armazenam valores diferentes de “1”, foi identificada a necessidade de adicioná-las nas dimensões. A razão para essa mudança é que adicionando-as nas medidas, não seria possível agrupar e contabilizar os valores armazenados nelas, sendo possível usar somente funções de agregação para cada combinação realizada. Para uma melhor compreensão desta mudança, foi elaborada a Tabela 6, exemplificando a necessidade. Onde para a medida “med\_qtd\_acid\_c\_lesoes\_ult\_12m”, cada entrevistado respondeu a quantidade de acidentes com lesões que sofreu nos últimos 12 meses.

Tabela - Medidas com valores diferentes de "1"

| Entrevistado | Valor |
| --- | --- |
| Entrevistado 1 | 2 |
| Entrevistado 2 | 1 |
| Entrevistado 3 | 2 |
| Entrevistado 4 | 1 |
| Entrevistado 5 | 1 |
| Entrevistado 6 | 2 |
|  |  |
| Soma dos valores | 10 |
| Média dos valores | 1,6 |
| Menor dos valores | 1 |
| Maior dos valores | 3 |
| Total de valores | 6 |
| Total de valores distintos | 3 |

Fonte: Autor, 2016.

Através desta mudança onde foram adicionadas as então medidas que armazenam valores diferentes de “1” para a área de dimensões do cubo, dentro do *Pentaho Community*, foi possível realizar o agrupamento e contagem dos valores armazenados, de acordo com a combinação das demais dimensões, gerando um melhor resultado para as análises. O resultado desta seleção de medidas, dimensões, atributos e nomes pode ser visualizado na Tabela 7.

Tabela - Atributos do cubo “acidentes”

| Nome | Tipo | *View* | Atributo |
| --- | --- | --- | --- |
| Qtd. entrevistados | Medida | fato\_acidentes | med\_contador |
| Unidade de federação/Nome | Dimensão | dim\_un\_fed | estado |
| Unidade de federação/Região | Dimensão | dim\_un\_fed | região |
| Sexo | Dimensão | dim\_sexo | descricao |
| Faixa etária | Dimensão | dim\_fa\_eta | descricao |
| Cor/Raça | Dimensão | dim\_cor\_raca | descricao |
| Estado civil | Dimensão | dim\_est\_civil | descricao |
| Sabem ler e escrever | Dimensão | dim\_sabe\_ler\_esc | descricao |
| Curso que frequentam | Dimensão | dim\_curso\_que\_freq | descricao |
| Possuem curso sup. Completo | Dimensão | dim\_possui\_curso\_sup\_comp | descricao |
| Dirigem carro | Dimensão | dim\_dir\_car | descricao |
| Dirigem motocicleta | Dimensão | dim\_dir\_mot | descricao |
| Freq. que andam de autom., van ou taxi | Dimensão | dim\_freq\_anda\_aut\_van\_taxi | descricao |
| Se envolveram em acid. c/lesões ult. 12m | Dimensão | dim\_envol\_acid\_c\_lesoes\_ult\_12m | descricao |
| Qtd. acid. c/lesões ult. 12m | Dimensão | fato\_acidentes | med\_qtd\_acid\_c\_lesoes\_ult\_12m |
| Ocupação no acid. mais grave ult. 12m | Dimensão | dim\_ocup\_veic\_acid\_mais\_grave | descricao |

Fonte: Autor, 2016.

### Cubo “uso\_alcool”

Para o desenvolvimento do cubo “uso\_alcool”, foram utilizadas as mesmas etapas que foram detalhadas e realizadas na seção anterior deste mesmo capítulo para o cubo acidentes, são elas:

* Seleção da *view* fato e suas *views* dimensões;
* Criação dos *joins* para cada uma das *views* dimensões com a *view* fato;
* Seleção e configuração dos atributos responsáveis por mostrar os valores nas medidas e dimensões do cubo, os quais podem ser melhor compreendidos observando-se a Tabela 8.

Tabela - Atributos do cubo “uso\_alcool”

| Nome | Tipo | View | Atributo |
| --- | --- | --- | --- |
| Qtd. entrevistados | Medida | fato\_uso\_alcool | med\_contador |
| Unidade de federação/Nome | Dimensão | dim\_un\_fed | estado |
| Unidade de federação/Região | Dimensão | dim\_un\_fed | regiao |
| Sexo | Dimensão | dim\_sexo | descricao |
| Faixa etária | Dimensão | dim\_fa\_eta | descricao |
| Cor/Raça | Dimensão | dim\_cor\_raca | descricao |
| Estado civil | Dimensão | dim\_est\_civil | descricao |
| Sabe ler e escrever | Dimensão | dim\_sabe\_ler\_esc | descricao |
| Curso que frequenta | Dimensão | dim\_curso\_que\_freq | descricao |
| Possui curso sup. completo | Dimensão | dim\_possui\_curso\_sup\_comp | descricao |
| Freq. que cost. consumir beb. alc. | Dimensão | dim\_freq\_costu\_consu\_beb\_alc | descricao |
| Dirigiram após consumir beb. alc. | Dimensão | dim\_dirigiu\_apos\_consu\_beb\_alc | descricao |
| Consumiram mais que 4 doses em uma unica ocasião no ult. mês | Dimensão | dim\_ult\_30d\_consu\_mais\_4\_dos\_beb\_alc\_un\_vez | descricao |
| Qtd. dias que consumiram mais que 4 doses em uma única vez no ult. mês | Dimensão | dim\_qt\_dia\_mes\_consu\_mais\_4\_dos\_beb\_alc\_un\_vez | descricao |
| Qtd. dias na semana que costumam consumir beb. alc. | Dimensão | fato\_uso\_alcool | med\_qtd\_dia\_sema\_costu\_consu\_beb\_alc |
| Qtd. doses por dia que costumam consumir de beb. alc. | Dimensão | fato\_uso\_alcool | med\_qtd\_dos\_dia |
| Idade que começou a cons. beb. alc. | Dimensão | fato\_uso\_alcool | med\_idade\_comec\_consu\_beb\_alc |

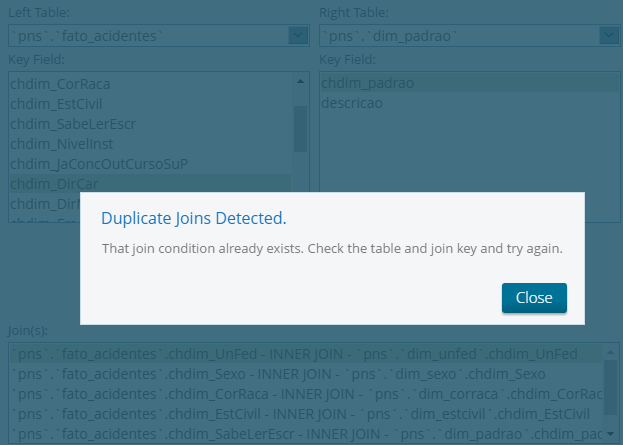
Fonte: Autor, 2016.

### Restrição de realizar um único *join* entre a fato e uma dimensão no *Pentaho Community*

Inicialmente na modelagem e criação das *views*, que representam as dimensões, foi criada uma nova *view* para representar uma dimensão padrão. Sua função seria substituir as dimensões que armazenam valores exatamente iguais em seus atributos. Mais especificamente para substituir as dimensões que armazenam valores “Sim”, “Não” e “Não aplicável”, como por exemplo, as dimensões: dim\_dir\_car; dim\_sabe\_ler\_esc; dim\_possui\_curso\_sup\_comp.

Porém foi identificado que o *Pentaho Community* não permite que seja criado mais do que um *join* com uma única dimensão. Ou seja, se a chave da *view* dimensão “dim\_padrao” já estiver ligada com a chave “ch\_dim\_sabe\_ler\_esc” da *view* fato “fato\_acidentes”, não será possível realizar outro *join* da *view* dimensão “dim\_padrao” com outra chave da *view* fato “fato\_acidentes”. A imagem exibida na Figura 29 demonstra essa restrição.

Figura - Restrição de realizar um único *join* entre a tabela fato e uma tabela dimensão no *Pentaho Community*



Fonte: Autor, 2016.

## Visualização OLAP e avaliação dos resultados

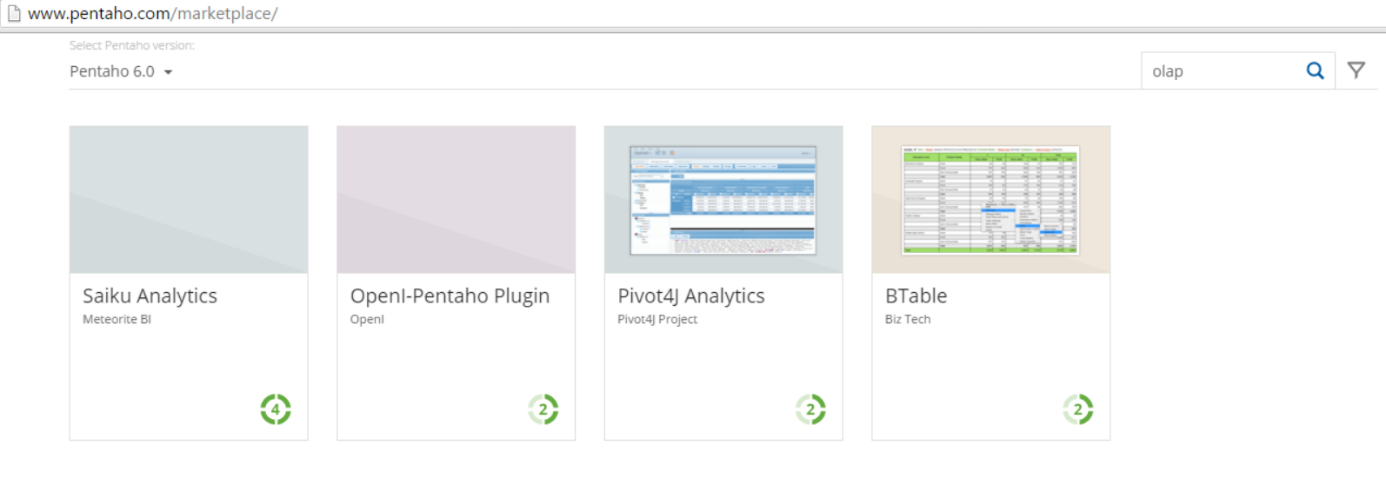
Com o objetivo de comparar dois *plug-ins* OLAP compatíveis com o ambiente do *Pentaho Community*, os cubos “acidentes” e “uso\_alcool” foram apresentados cada um utilizando um *plug-in* diferente. Para o cubo de acidentes foi utilizado o *plug-in* *jPivot*, o qual já veio pronto para ser utilizado após a conclusão da instalação do ambiente efetuada na seção 5.1 deste capítulo. Já para o cubo, “uso\_alcool”, foi realizada uma busca de *plug-ins* adicionais com recursos OLAP dentro do site do *Pentaho Community*.

Para realizar esta busca foi utilizada a expressão “OLAP”. Dentre os resultados encontrados, foi necessário definir um critério para a escolha de um, de modo que foi utilizado a classificação do próprio site, que os classifica de acordo com o estágio de desenvolvimento em que se encontram, são eles:

* ***Development* (1)**: Fase de desenvolvimento. São altamente instáveis;
* ***Snapshot* (2)**: Versão instável provavelmente, pode ser utilizado para testes e avaliações, mas não é recomendado para uso em produção;
* ***Stable* (3)**: Versão estável de lançamento. Já pode ser utilizado em ambientes de produção;
* ***Mature* (4)**: Última etapa de um projeto, indica que ele foi aprovado com sucesso durante e após a etapa de lançamento (HITACHI, 2015).

Dentre os resultados encontrados, para a visualização e apresentação do cubo “uso\_alcool”, foi optado pelo *Saiku Analytics*. Único *plug-in* que se encontrava no último estágio “*Mature*”, conforme é demonstrado pela Figura 30.

Figura - *Plug-ins* adicionais de OLAP para o *Pentaho Community*

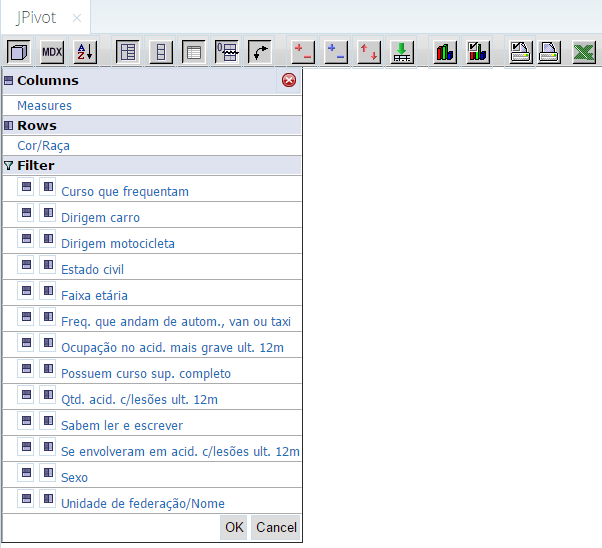


Fonte: Autor, 2016.

### Visualização OLAP do cubo “acidentes”

Com o uso do *plug-in* OLAP *jPivot* foi possível realizar a leitura dos cubos criados a partir do DW responsável por armazenar os dados da PNS 2013. Através dele foi criada uma consulta demonstrando algumas de suas funcionalidades e recursos. Dentre todas as dimensões disponíveis para o cubo “acidentes”, conforme demonstra a Figura 31, foi construída uma visualização utilizando as dimensões “Se envolveram em acid. c/lesões ult. 12m”, “Sexo” e “Estado civil”. Para a métrica foi utilizado a medida “Qtd. entrevistados”.

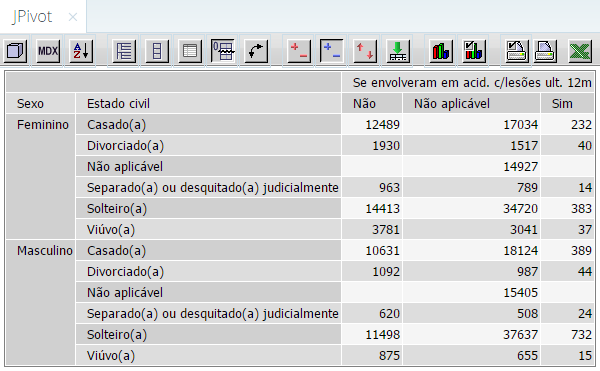
Figura – Todas as dimensões e medidas do cubo “acidentes”



Fonte: Autor, 2016.

Na área *Columns* do cubo foi adicionada a medida “Qtd. entrevistados”, juntamente com a dimensão “Se envolveram em acid. c/lesões ult. 12m”. Já para a área *Rows* foram incluídas as dimensões “Sexo” e “Estado civil”. O resultado da visualização para esta consulta é demonstrado pela Figura 32.

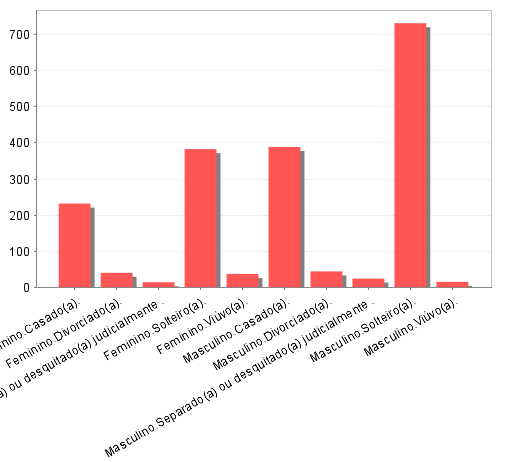
Figura - Visualização OLAP do cubo “acidentes” (“Qtd. entrevistados” x “Se envolveram em acid. c/lesões ult. 12m” x “Sexo” x “Estado civil”)



Fonte: Autor, 2016.

Para a consulta realizada com o cubo “acidentes”, ilustrada pela Figura 32, é possível se chegar a algumas conclusões. A principal delas mostra que a maioria dos entrevistados que responderam “Sim” para a pergunta armazenada na dimensão “Se envolveram em acid. c/lesões ult. 12m” é do sexo masculino. Além disto, analisando a dimensão “Estado civil”, também é possível concluir que estes são solteiros, o que pode ser comprovado observando o gráfico gerado pela própria ferramenta, o qual é demonstrado pela Figura 33.

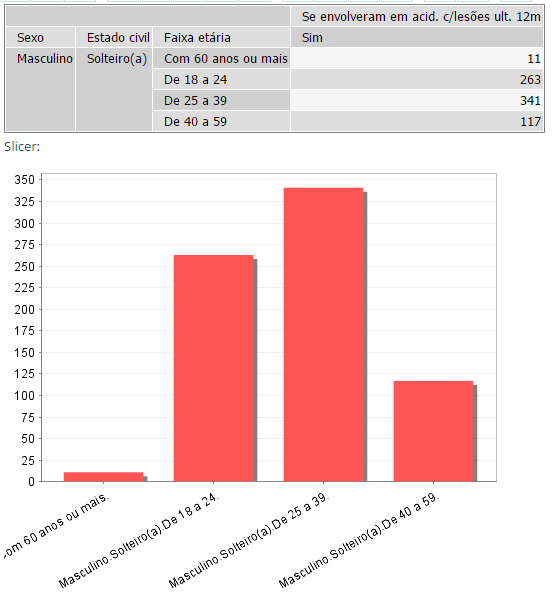
Figura – Gráfico OLAP do cubo “acidentes” (“Qtd. entrevistados” x “Se envolveram em acid. c/lesões ult. 12m” = sim x “Sexo” x “Estado civil”)



Fonte: Autor, 2016.

Com o intuito de deixar a consulta ainda mais detalhada e identificar qual a faixa etária dos homens solteiros que mais se envolveram em acidentes com lesões nos últimos 12 meses e também demonstrar o recurso de *drill-down*, foi adicionada ao cubo à dimensão “Faixa etária” na área *Rows,* logo após a dimensão “Estado civil”. A partir da análise do resultado da consulta, o qual é ilustrado pela Figura 34, é possível concluir que: para os entrevistados da PNS 2013 que se envolveram em acidentes com lesões nos últimos 12 meses, a maioria são do sexo masculino, solteiros e possuem entre 25 e 39 anos de idade.

Figura - Visualização e gráfico OLAP do cubo “acidentes” (“Qtd. entrevistados” x “Se envolveram em acid. c/lesões ult. 12m” x “Sexo” x “Estado civil” x “Faixa etária”)

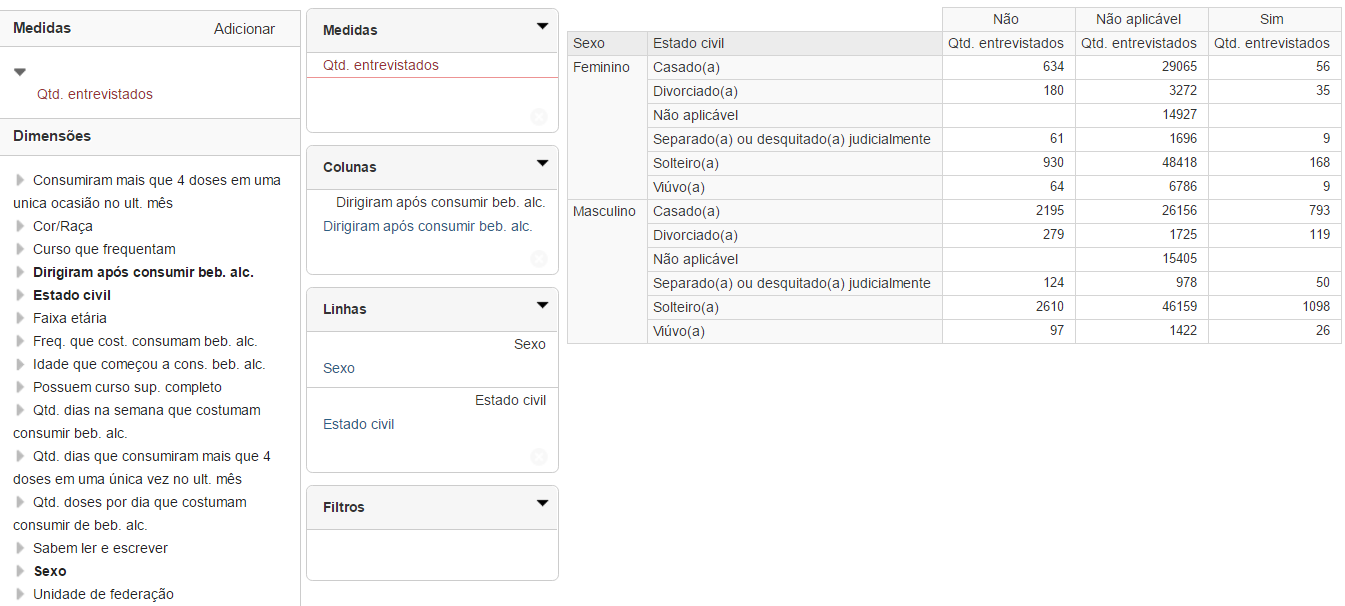


Fonte: Autor, 2016.

### Visualização OLAP do cubo “uso\_alcool”

Através do *plug-in* OLAP *Saiku Analytics*, assim como foi feito na seção anterior com o *plug-in jPivot*, foi desenvolvida uma visão utilizando as medidas e dimensões do cubo “uso\_alcool”. Para esta análise foram utilizadas as dimensões “Dirigiram após consumir beb. alc.”, “Sexo” e “Estado civil”. Para visualizar as métricas foi adicionada a medida “Qtd. entrevistados”. A Figura 35 demonstra o resultado desta visualização.

Figura - Visualização OLAP do cubo “uso\_alcool” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” x “Sexo” x “Estado civil”)



Fonte: Autor, 2016.

Além do resultado da consulta realizada, observando a Figura 35 é possível identificar em sua esquerda todas as dimensões e medidas disponíveis do cubo “uso\_alcool” para a construção das análises. Para a visualização do cubo em questão, foi utilizada como métrica a medida “Qtd. entrevistados” juntamente com a dimensão “Dirigiram após consumir beb. alc.” na área Colunas e das dimensões “Sexo” e “Estado civil” na área Linhas. A partir do resultado obtido pode-se chegar a conclusão que: a maioria dos entrevistados da PNS 2013 que já dirigiram após consumir bebidas alcoólicas são do sexo masculino e solteiros.

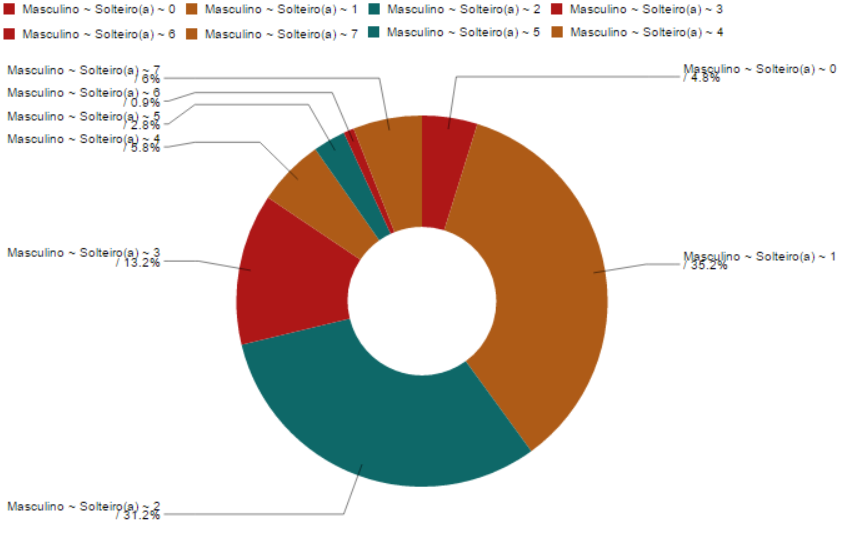
Utilizando os filtros da ferramenta, foram ocultados os demais detalhes que não se encaixavam no resultado obtido na consulta anterior e adicionada à dimensão “Qtd. doses por dia que costumam consumir de beb. alc.”, para demonstrar o recurso de *drill-down*. O resultado desta nova análise pode ser visualizado observando-se a Figura 36 e também o gráfico da Figura 37, onde foi possível identificar que a maioria dos homens solteiros que já dirigiram após consumir bebida alcoólica, fazem isso um dia na semana.

Figura - Visualização OLAP do cubo “uso\_alcool” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” x “Sexo” x “Estado civil” x “Qtd. dias na semana que costumam consumir beb. alc.”)



Fonte: Autor, 2016.

Figura – Gráfico do cubo “uso\_alcool” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” x “Sexo” x “Estado civil” x “Qtd. dias na semana que costumam consumir beb. alc.”)



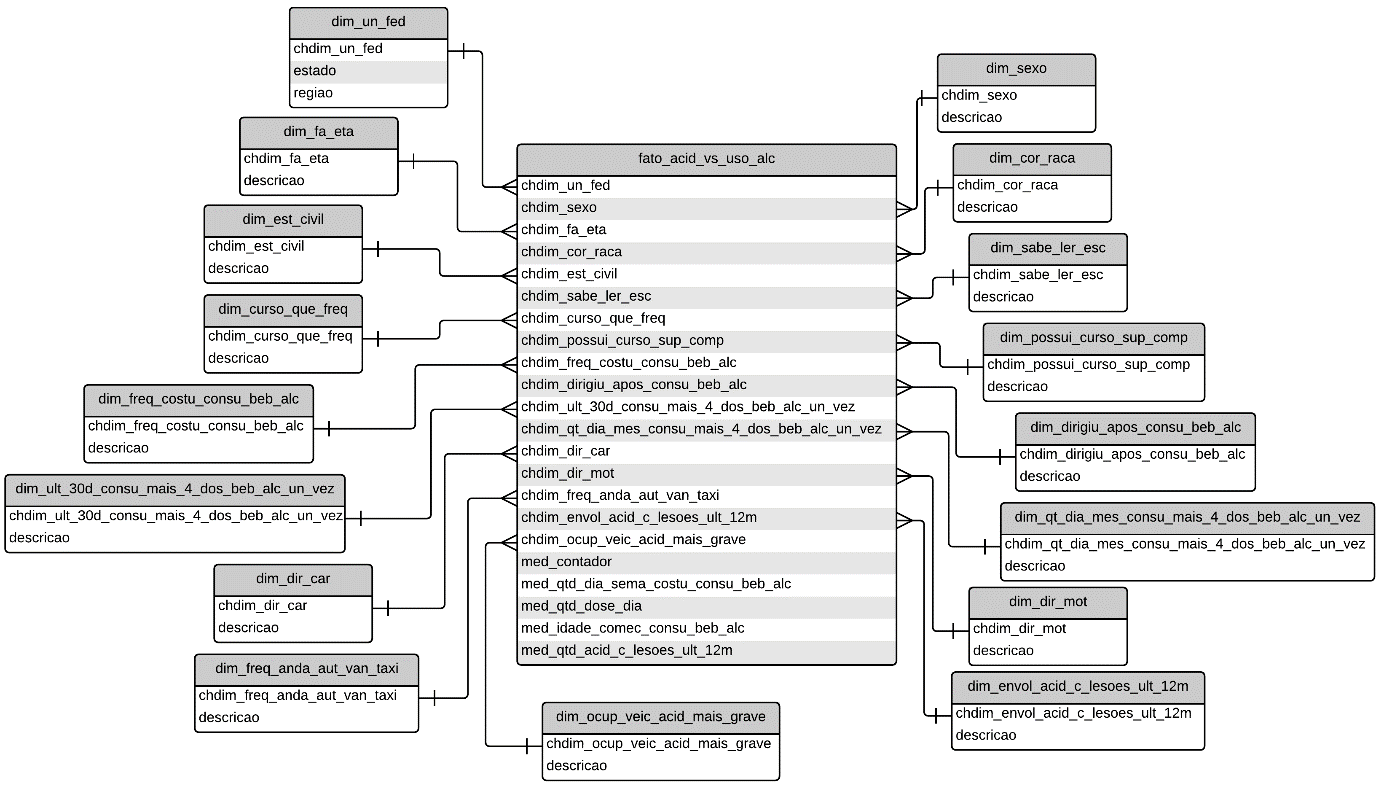
Fonte: Autor, 2016.

De acordo como foi citado no início da seção 5.4 deste capítulo, com o objetivo de comparar dois *plug-ins* OLAP, foi definido que para cada um dos dois cubos existentes, fosse utilizado um *plug-in* OLAP diferente. Para o cubo “acidentes” o *plug-in jPivot* e para o cubo “uso\_alcool” o *plug-in* *Saiku*. Como resultado dessa comparação, pode-se se concluir que ambos exibem com clareza as informações das análises e visões criadas. Porém no momento da criação destas análises e visões, foi observado diversas facilidades do *plug-in* *Saiku* em relação ao *jPivot*. Sendo uma das principais o recurso que permite arrastar e soltar os atributos das dimensões e medidas direto para a área onde é exibida a análise. Outro ponto positivo observado no *Saiku,* foi em relação a sua interface e apresentação dos dados, tendo ele um visual mais moderno e intuitivo em relação ao *jPivot*.

### Criação e visualização OLAP do cubo “acid\_vs\_uso\_alc”: cruzamento dos cubos “acidentes” e “uso\_alcool”

Nesta seção foi desenvolvido um novo cubo, que utilizou e cruzou os dados dos já existentes “acidentes” e “uso\_alcool”. De modo que primeiramente foi preciso criar uma nova *view*, responsável por representar o fato “acid\_vs\_uso\_alc”, o qual é ilustrado pela Figura 38. Seu papel é armazenar os atributos em comum e distintos das *views* “fato\_acidentes” e “fato\_uso\_alcool”, as quais foram criadas nas seções 4.2.1 e 4.2.2 do capítulo 4.

Figura - Modelo estrela da *view* “fato\_acid\_vs\_uso\_alc” e suas dimensões



Fonte: Autor, 2016.

De posse desta nova *view* fato “fato\_acid\_vs\_uso\_alc”, ilustrada pela Figura 38, foi possível realizar a construção e configuração do cubo dentro do ambiente do *Pentaho Community*. Para isso foram utilizadas as mesmas etapas já descritas na seção 5.3, são elas: Seleção da *view* fato e suas *views* dimensões; Criação dos *joins* para cada uma das *views* dimensões com a *view* fato; Seleção e configuração dos atributos responsáveis por mostrar os valores nas medidas e dimensões do cubo, demonstrados na Tabela 9.

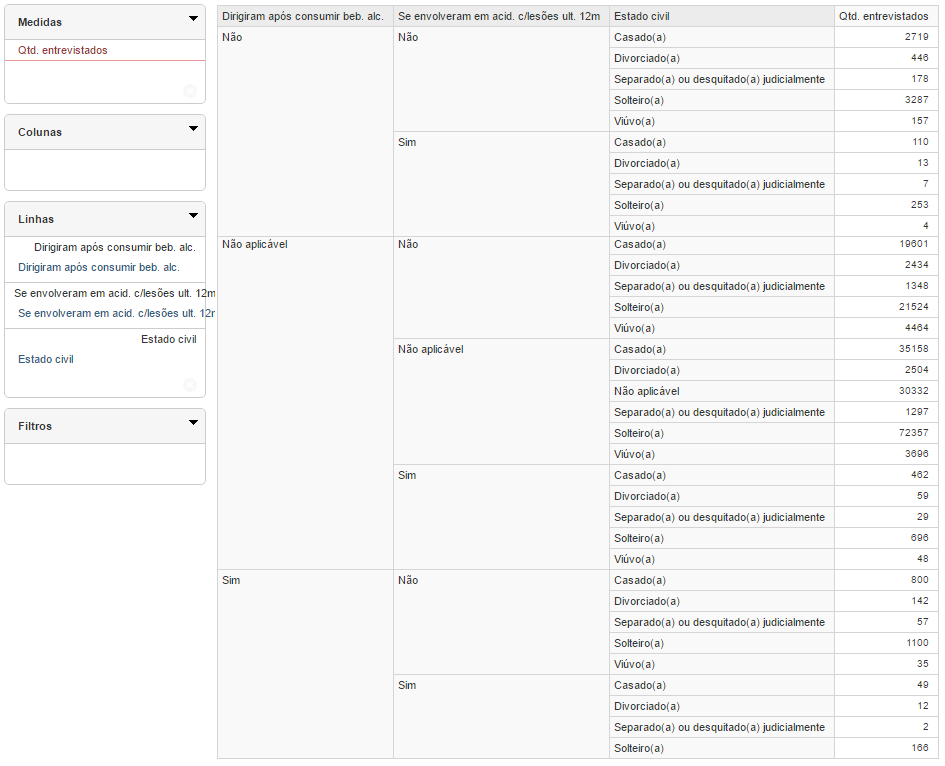
Tabela – Atributos do cubo “acid\_vs\_uso\_alc”

| Nome | Cubo | Tipo | View | Atributo |
| --- | --- | --- | --- | --- |
| Qtd. entrevistados | Em comum | Medida | fato\_acidentes | med\_contador |
| Unidade de federação/Nome | Em comum | Dimensão | dim\_un\_fed | estado |
| Unidade de federação/Região | Em comum | Dimensão | dim\_un\_fed | regiao |
| Sexo | Em comum | Dimensão | dim\_sexo | descricao |
| Faixa etária | Em comum | Dimensão | dim\_fa\_eta | descricao |
| Cor/Raça | Em comum | Dimensão | dim\_cor\_raca | descricao |
| Estado civil | Em comum | Dimensão | dim\_est\_civil | descricao |
| Sabem ler e escrever | Em comum | Dimensão | dim\_sabe\_ler\_esc | descricao |
| Curso que frequentam | Em comum | Dimensão | dim\_curso\_que\_freq | descricao |
| Possuem curso sup. Completo | Em comum | Dimensão | dim\_possui\_curso\_sup\_comp | descricao |
| Dirigem carro | acidentes | Dimensão | dim\_dir\_car | descricao |
| Dirigem motocicleta | acidentes | Dimensão | dim\_dir\_mot | descricao |
| Freq. que andam de autom., van ou taxi | acidentes | Dimensão | dim\_freq\_anda\_aut\_van\_taxi | descricao |
| Se envolveram em acid. c/lesões ult. 12m | acidentes | Dimensão | dim\_envol\_acid\_c\_lesoes\_ult\_12m | descricao |
| Qtd. acid. c/lesões ult. 12m | acidentes | Dimensão | fato\_acidentes | med\_qtd\_acid\_c\_lesoes\_ult\_12m |
| Ocupação no acid. mais grave ult. 12m | acidentes | Dimensão | dim\_ocup\_veic\_acid\_mais\_grave | descricao |
| Freq. que cost. consumir beb. alc. | uso\_alcool | Dimensão | dim\_freq\_costu\_consu\_beb\_alc | descricao |
| Dirigiram após consumir beb. alc. | uso\_alcool | Dimensão | dim\_dirigiu\_apos\_consu\_beb\_alc | descricao |
| Consumiram mais que 4 doses em uma unica ocasião no ult. Mês | uso\_alcool | Dimensão | dim\_ult\_30d\_consu\_mais\_4\_dos\_beb\_alc\_un\_vez | descricao |
| Qtd. dias que consumiram mais que 4 doses em uma única vez no ult. Mês | uso\_alcool | Dimensão | dim\_qt\_dia\_mes\_consu\_mais\_4\_dos\_beb\_alc\_un\_vez | descricao |
| Qtd. dias na semana que costumam consumir beb. alc. | uso\_alcool | Dimensão | fato\_uso\_alcool | med\_qtd\_dia\_sema\_costu\_consu\_beb\_alc |
| Qtd. doses por dia que costumam consumir de beb. alc. | uso\_alcool | Dimensão | fato\_uso\_alcool | med\_qtd\_dos\_dia |
| Idade que começou a cons. beb. alc. | uso\_alcool | Dimensão | fato\_uso\_alcool | med\_idade\_comec\_consu\_beb\_alc |

Fonte: Autor, 2016.

Após a criação do cubo, através do *plug-in* OLAP *Saiku* foi criada uma nova visualização, a qual exibe os dados originados do cruzamento entre os cubos “acidentes” e “uso\_alcool”. Para esta análise foi utilizada a medida “Qtd. entrevistados” e as dimensões: “Dirigiram após consumir beb. alc., do cubo “uso\_alcool”; “Se envolveram em acid. c/lesões ult. 12m”, do cubo “acidentes”; e “Estado civil”, comum entre os cubos. O resultado desta análise é ilustrado pela Figura 39.

Figura - Visualização OLAP do cubo “acid\_vs\_uso\_alc” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” x “Se envolveram em acid. c/lesões ult. 12m” x “Estado civil”)

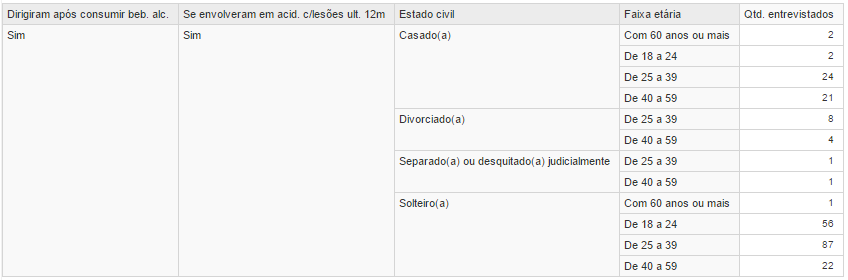


Fonte: Autor, 2016.

Através da análise realizada com o cubo “acid\_vs\_uso\_alc”, ilustrada pela Figura 39, foi possível chegar a conclusão de que a grande maioria dos entrevistados da PNS 2013 que responderam “Sim” para as perguntas “dirigiram após consumir álcool” e “se envolveram em acidentes com lesões nos últimos 12 meses” são solteiros.

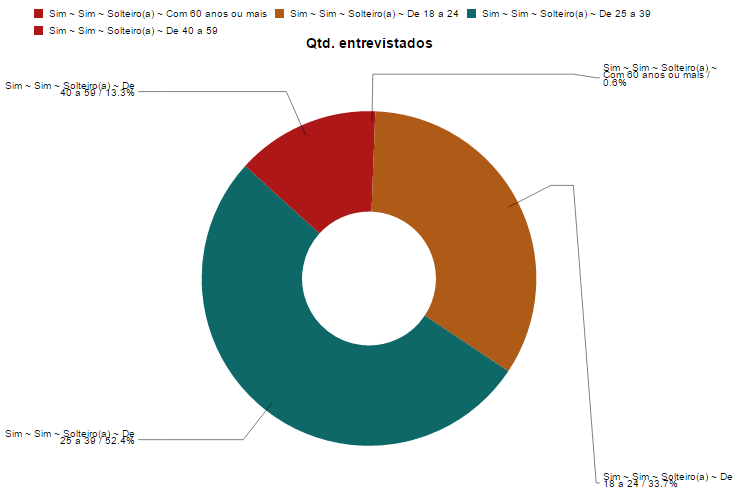
Para deixar a visualização mais clara e objetiva, através dos filtros da ferramenta, foram ocultados os resultados que não se encaixavam na conclusão chegada anteriormente. Além disso, com o objetivo de adicionar mais detalhes a consulta, também foi incluída a dimensão “Faixa etária”, a qual é comum entre os dois cubos. Através deste complemento na análise, a qual é ilustrada pela Figura 40 e pelo gráfico demonstrado na Figura 41, foi possível concluir que dos entrevistados da PNS 2013 que responderam “Sim” para as perguntas “dirigiram após consumir álcool” e “se envolveram em acidentes com lesões nos últimos 12 meses” são solteiros e possuem de 25 a 39 anos de idade.

Figura - Visualização OLAP do cubo “acid\_vs\_uso\_alc” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” = “Sim” x “Se envolveram em acid. c/lesões ult. 12m” = “Sim” x “Estado civil” x “Faixa etária”)



Fonte: Autor, 2016.

Figura - Gráfico do cubo “acid\_vs\_uso\_alc” (“Qtd. entrevistados” x “Dirigiram após consumir beb. alc.” = “Sim” x “Se envolveram em acid. c/lesões ult. 12m” = “Sim” x “Estado civil” = “Solteiro” x “Faixa etária”)



Fonte: Autor, 2016.

Com base nos resultados obtidos através das análises e visões realizadas, foi possível identificar que todas elas são de grande importância e de interesse da maioria da população nacional. Além de possuírem um grande potencial para serem discutidas por profissionais da área da saúde e também pelos governos municipais, estaduais e federal. Dando a eles a possibilidade de criar novas consultas e análises a partir dos cubos “acidentes”, “uso\_alcool” e “acid\_vs\_uso\_alc”.

## Entrevistas

A fim de extrair uma percepção sobre o ambiente de BI criado neste trabalho e dos dados os quais ele armazena, foram escolhidos e entrevistados três profissionais com conhecimento e interesse em saúde e *Business Intelligence*. Para cada um deles, foram estipulados alguns requisitos, os quais deveriam atender. São eles:

* Entrevistada 1: trabalhar na área da saúde e possuir experiência com BI;
* Entrevistada 2: trabalhar e ter experiência na área da saúde;
* Entrevistado 3: trabalhar na área de sistemas de informação e ter experiência com mais do que um sistema de BI.

A primeira entrevistada foi uma Assessora Administrativa de um grande hospital Federal situado na cidade de Porto Alegre/RS, Administradora, Especialista em Gestão em Saúde e Mestre em Epidemiologia. Ela possui mais de 10 anos de experiência com sistemas de BI voltados a saúde. Já a segunda entrevistada foi com uma Enfermeira e Professora de Ensino Superior – licenciada para realizar doutorado sanduíche no exterior com amplo conhecimento na área da saúde, porém sem conhecimento em BI. O terceiro e último entrevistado, foi um empresário do ramo de tecnologia, Analista de Sistemas com conhecimento em sistemas de BI e também no Pentaho.

A primeira etapa da entrevista consistiu em realizar uma avaliação sobre a usabilidade do *Pentaho Community* e seus dois *plug-ins* OLAP *jPivot* e *Saiku Analyctis*. Durante esta etapa também foi apresentada a visualização dos três cubos criados: “acidentes”; “uso\_alcool”; e “acid\_vs\_uso”. De maneira que cada entrevistado deu seu parecer sobre a percepção das informações contidas em cada um dos cubos, a quem eles podem beneficiar e quais os dados podem ser agregados aos cubos. Por fim, foi solicitado que cada um dos entrevistados desse um parecer sobre o ambiente de BI e realizasse uma comparação com outras ferramentas.

Com relação a usabilidade dos dois *plug-ins* OLAP apresentados, todos comentaram que ambos atendem bem ao propósito final, porém foram unânimes em relação ao *plug-in* *Saiku Analyctis*, destacando sua interface amigável e facilidade de uso e navegação. No que diz respeito das informações contidas nos cubos, a “entrevistada 1” observou que, se ampliados, os dados serão úteis tanto para diagnóstico, quanto para definição de intervenção. Possibilitando que gestores possam, a partir do cruzamento de informações da PNS, melhorar o planejamento e a priorização de recursos visando resultados mais efetivos na redução dos acidentes de trânsito e dos demais danos decorrentes destes. Conhecer o perfil do motorista permitirá, por exemplo, ações de educação mais eficazes para a prevenção e redução de acidentes com mortes e que resultam em incapacidade.

Já a “entrevistada 2” ressaltou a importância dos dados de domínio público estarem disponíveis no DATASUS, os quais possibilitam o desenvolvimento de ferramentas que facilitem as análises dos seus dados. Sobre as informações contidas nos cubos, destaca as múltiplas possibilidades de cruzamento que fornecem maior riqueza às informações, as quais poderão ser utilizadas por gestores e pesquisadores para subsidiar a tomada de decisões, no planejamento de novas pesquisas ou na destinação de recursos públicos. Como sugestão, acrescentou que poderia ser agregado aos cubos outros dados disponíveis no DATASUS, como a prevalência de doenças que tem como fator de risco o abuso de álcool.

Já o “entrevistado 3” destacou a importância dos órgãos ligados a saúde e segurança possuírem sistemas de BI, reforçou também a importância dos cubos criados e das informações que podem ser obtidas através deles. Observou também que eles poderão ajudar estes órgãos nos processos de tomadas de decisões, como por exemplo criar e direcionar campanhas de conscientização referente ao consumo de álcool relacionado com acidentes.

Para a última parte da entrevista, os entrevistados “1” e “3”, observaram algumas semelhanças do Pentaho com os outros ambientes de BI que já possuem conhecimento e experiência. Dentre elas destacaram a facilidade de manipular os cubos, sendo eles bem intuitivos e fáceis de navegar. Por fim, destacaram que o Pentaho pode ser uma excelente alternativa, pois é oferecido de forma gratuita e possui uma ampla comunidade ativa de desenvolvedores e utilizadores.

# CONCLUSÃO

A partir das pesquisas que foram realizadas foi possível observar que os sistemas de *Business Intelligence* oferecem diversos recursos importantes e diferenciados para uma organização. Com o intuito de aumentar a sua competitividade em um mercado que está cada vez mais concorrido, estas organizações estão optando pelo uso de ferramentas de *Business Intelligence* como um diferencial para obter vantagem frente aos seus concorrentes. Pois através do seu uso é possível realizar a integração de diversas fontes de dados em um só ambiente, disponibilizando de forma rápida e segura informações confiáveis para auxiliar seus gestores nas tomadas de decisões e definir novas estratégias de mercado.

Durante a implementação do ambiente de *Business Intelligence*, para utilizar suas ferramentas, conceitos e técnicas, foi necessário dispor de um grande volume de dados para submetê-los a estas ferramentas. Para isso foi identificado que o portal do Departamento de Informática do SUS (DATASUS) dispõe de diversos repositórios de dados que estão organizados e divididos em diversas áreas da saúde brasileira. Dados que ainda não foram explorados por sistemas de *Business Intelligence* e que estão disponíveis em arquivos organizados de maneira bidimensional, ou seja, linhas e colunas, gerando um certo grau de dificuldade para realizar cruzamentos entre as consultas e também gerar relatórios mais elaborados e dinâmicos.

Com base nos conhecimentos adquiridos no decorrer deste trabalho, foi implementado uma solução de *Business Intelligence* utilizando parte do grande volume de dados disponíveis no portal do DATASUS. Com o auxílio de técnicas e ferramentas de *ETL*, os dados foram extraídos, transformados e carregados para um *Data Warehouse*, onde foi possível analisá-los através de ferramentas *OLAP*.

Também é importante destacar todo o conhecimento adquirido no ambiente de BI do Pentaho durante a realização deste trabalho. Uma vez que para sua apresentação, foi necessário realizar todo o desenvolvimento do ambiente, desde criação dos processos de ETL, passando pela construção do DW e dos cubos até o instalação e manipulação do Pentaho.

Além de todo o conhecimento obtido durante a pesquisa e elaboração deste trabalho, entende-se que através do uso destas técnicas e ferramentas de *Business Intelligence*, foi possível analisar de forma mais clara, objetiva e dinâmica uma fatia do grande volume de dados disponível no portal do DATASUS. Isso permite aos profissionais da área da saúde e também a população brasileira a criação de consultas *OLAP*, cruzamentos e análises multidimensionais a partir dos dados da PNS 2013 que foram extraídos do portal do DATASUS. Além dos dados que foram extraídos, existe uma grande variedade de outros dados, os quais podem ser utilizados para desenvolver novos ambientes de BI.

Também destaca-se que todo o processo de desenvolvimento deste trabalho contribuiu para avançar no conhecimento enquanto profissional de Sistemas de Informação, uma vez que permitiu um estudo aprofundado e a aplicação dos conteúdos das diversas disciplinas do curso. Ainda, salienta-se a importância deste trabalho que possibilita a análise integrada de informações disponíveis no DATASUS. Estas informações são imprescindíveis para a elaboração de políticas públicas e de estratégias para o enfrentamento de agravos relevantes na saúde pública, ou seja, é fundamental conhecer e visualizar os dados a partir destas análises para criar as estratégias.

Também, através das entrevistas realizadas, foi possível comprovar a importância dos dados extraídos da PNS de 2013, os quais puderam mostrar diversas visualizações relevantes a respeito do uso de álcool e acidentes de trânsito. Tendo também como base os cubos desenvolvidos neste trabalho e as informações geradas através deles, acredita-se que eles possam ser de grande utilidade se usados por órgãos de saúde e de trânsito.

Por fim, foi identificado três opções para trabalhos futuros. A primeira delas, é desenvolver novos cubos a partir dos módulos da PNS de 2013 que não foram explorados neste trabalho. Já a segunda opção, é disponibilizar o ambiente de BI desenvolvido em algum meio público, para que toda a população brasileira tenha acesso a ele de forma gratuita. E como última opção, explorar e adicionar os dados das PNS anteriores a 2013, com a finalidade de montar cubos que permitam uma análise em relação ao tempo.

# Referências Bibliográficas

BARBIERI, Carlos. ***Business Intelligence*: Modelagem & Tecnologia**. Rio de Janeiro: Axcel Books, 2001. 424 p.

DATASUS A. **Departamento de Informática do SUS**. Disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=02>. Acesso em: mai. 2015.

DATASUS B. **Departamento de Informática do SUS**. Disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=04>. Acesso em: mai. 2015.

DATASUS C. **Departamento de Informática do SUS**. Disponível em <http://tabnet.datasus.gov.br/cgi/idb2000/fqc12.htm>. Acesso em: mai. 2015.

DATASUS D. **Departamento de Informática do SUS**. Disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=0207&id=28247790>. Acesso em: mai. 2015.

ELMASRI, Ramez; NAVATHE, B. Shamkant. **Sistemas de banco de dados**. São Paulo: *Pearson Education* do Brasil, 2011. 788 p.

HITACHI, *Data Systems Company*. **Pentaho**. Disponível em <http://www.pentaho.com/about/>. Acesso em: ago. 2015.

IBGE. **Instituto Brasileiro de Geografia e Estatística**. Disponível em <http://www.ibge.gov.br/home/estatistica/populacao/pns/2013/default.shtm>. Acesso em: abr. 2016.

IBM. ***The IBM Corporate***. Disponível em < http://www-03.ibm.com/software/products/pt/cognos-analytics >. Acesso em: nov. 2015.

INMON, W. H.; TERDEMAN, R.H.; IMHOFF Claudia. ***Data* *warehousing*: Como transformar informações em oportunidades de negócios**. São Paulo, Berkeley, 2001. 265 p.

JACOBSON, Reed; MISNER, Stacia. **Microsoft SQL *Server* 2005 *Analysis Services*: Passo a passo**. Porto Alegre, Bookman, 2007. 352 p.

KAPLAN, Robert; NORTON, David. **Alinhamento: usando o *balanced scorecard* para criar sinergias corporativas**. Rio de Janeiro: Elsevier, 2006. 360 p.

KIMBALL, Ralph; ROSS, Margy. ***Data Warehouse Toolkit Third Edition***. Indianapolis, Indiana: John Wiley & Sons, 2013, 564 p.

MACHADO, Felipe N. R. **Projeto de *Data Wahehouse*: Uma Visão multidimensional**. São Paulo: Érica, 2000, 248 p.

MASCARENHAS, Marcio Denis Medeiros; BARROS, Marílisa Berti de Azevedo. **Evolução das internações hospitalares por causas externas nos sistema público de saúde – Brasil, 2002 a 2011. Epidemiologia e Serviços de Saúde**. Brasília, v. 24, n. 1, p. 19-29, abr. 2015.

MICROSOFT. **Microsoft *Corporation: Analysis Services***. Disponível em <https://msdn.microsoft.com/pt-br/library/bb522607.aspx>. Acesso em: dez 2015.

ORACLE. **Oracle *Corporation:* BI**. Disponível em <https://www.oracle.com/solutions/business-analytics/business-intelligence/index.html>. Acesso em: nov. de 2015.

TURBAN, Efraim; SHARDA, Ramesh; ARONSON, Jay; KING, David. ***Business Intelligence*: Um enfoque gerencial para a inteligência do negócio**. Porto Alegre: Artmed Editora, 2009. 253 p.

WHO. ***World Health Organization: Global status report on alcohol and health* 2014**. Disponível em <http://apps.who.int/iris/bitstream/10665/112736/1/9789240692763\_eng.pdf>. Acesso em: mar. 2016.

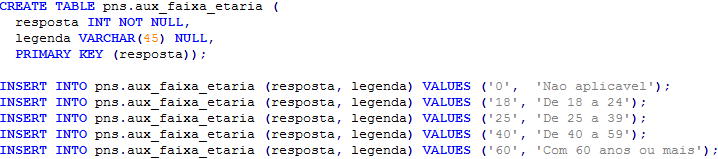
# ANEXO A - Tabela de Projetos de Business Intelligence desenvolvidos através do DATASUS

| Projeto | Descrição | Situação |
| --- | --- | --- |
| BI GAL | Disponibilização das informações geradas no sistema GAL sob a forma de painéis de indicadores (*dashboards*) e de relatórios. | Construção |
| PLATAFORMA BRASIL | Gestão e acompanhamento dos pareceres nos CEPs e CONEPs. | Construção |
| AUDITORIA | Gestão e Acompanhamento das Ações de Controle de Auditoria, por parte do Secretário da SGEP. | Construção |
| SISPRENATAL | Gestão e acompanhamento do Programa de Monitoramento e Avaliação do Pré-Natal, Parto, Puerpério e Criança. | Construção |
| COAGULOPATIA | Gestão do Perfil das Coagulopatias Hereditárias no Brasil. | Construção |
| SISCAN | Gestão e Acompanhamento do Programa de Câncer de Mama e de Colo de Útero. | Construção |
| BI SCNES | Disponibilizar sob a forma de painéis (*dashboards*), mapas e relatórios informações que possibilite realizar análises de tendências e situacionais sobre a rede nacional de estabelecimentos e profissionais de saúde que compõem o Cadastro Nacional de Estabelecimentos de Saúde-CNES. | Construção |
| BI COAGULOPATIAS E HEMOGLOBINOPATIAS | Disponibilizar sob a forma de painéis com indicadores (*dashboards*) e relatórios para auxilio a tomada de decisão sobre o controle e administração dos pacientes com doenças hemorrágicas decorrentes da deficiência quantitativa e/ou qualitativa de um ou mais fatores de coagulação e dos pacientes com alterações genéticas caracterizadas pela predominância da hemoglobina (Hb) S, sendo entre as doenças falciformes mais frequentes a anemia falciforme (HbSS), a S/Beta Talassemia (S/ß Tal.), as doenças SC, SD, SE (duplos heterozigotos) e outras mais raras, as quais fazem parte das hemoglobinopatias. | Construção |
| Produção SIA-SIH | Gestão e acompanhamento da Produção dos sistemas SIA e SIH. | Construção |
| SISGREF BI | Disponibilização das informações geradas no sistema SISGREF sob a forma de painéis de indicadores (*dashboards*) e de relatórios. | Homologação |
| BI e SUS Hospitalar | Desenvolvimento de ferramenta que permita ao DGH a análise de indicadores de saúde a partir dos dados extraídos das bases de dados do e-SUS Hospitalar. | Produção |
| Módulo gráfico para Tabnet | Dotar o software de tabulações do DATASUS na web (Tabnet) de um módulo de geração de gráficos do tipo barra, pizza, linha e volume. | Produção |
| DRAC Teto | Prover o DRAC de uma ferramenta construída com as técnicas de BI para apresentação das informações das apurações de AIHs pagas/rejeitadas, por competência, CNES e Gestor, com valor e quantidade. | Produção |
| DRAC MAC | Prover o DRAC de uma ferramenta construída com as técnicas de BI para apresentação das informações dos valores gastos com média complexidade, por município e esfera administrativa, obtidos nas bases de dados dos sistemas SIH e SIA. | Produção |
| HÓRUS (Básico e Estratégico) | Gestão de Estoque, Saída e Estoque de Medicamentos da Farmácia Básica e dos Medicamentos dos Programas de Saúde (diabete, hipertensão, etc) e Dispensação nas Farmácias Físicas e Monitoramento da Adesão ao Sistema Hórus. | Produção |
| SISREG | Geração e emissão de relatórios padronizados e ad hoc do sistema de regulação. | Produção |
| ESPELHO AIH | Emissão do espelho de AIH para DRAC e DENASUS. | Produção |
| MALÁRIA | Gestão e acompanhamento das Notificações de casos de Malária. | Produção |
| ESPELHO APAC | Emissão do espelho de APACS para DRAC e DENASUS. | Produção |
| HÓRUS ESPECIALIZADO | Gestão de Estoque, Saída e Estoque de Medicamentos de Alto Custo e Dispensação nas Farmácias Físicas. | Produção |
| MAIS MÉDICOS | Gestão dos Participantes do Programa Mais Médicos. | Produção |
| OUVIDOR SUS | Gestão das atividades das Ouvidorias do SUS. | Produção |
| SOS EMERGÊNCIA | Gestão dos tempos de atendimentos nos Hospitais vinculados ao programa S.O.S Emergências. | Produção |
| PERIÓDICOS | Gestão e acompanhamento dos acessos às publicações técnicas por trabalhadores da saúde. | Produção |
| FARMÁCIA POPULAR | Gestão e acompanhamento da distribuição de medicamentos do Programa Farmácia Popular. | Produção |
| SISMAT | Controle do almoxarifado Central do MS (Estoque, Financeiro, Validade). | Produção |
| INSUMOS ESTRATÉGICOS | Visão centralizada dos sistemas SISMAT, SIES e SICLOM (AIDS). | Produção |
| SIASI | Gestão do Programa de Acompanhamento da Saúde dos povos indígenas. | Produção |
| SIES | Gestão de distribuição para Estados e Municípios, a partir do almoxarifado Central do MS, dos Insumos Estratégicos. | Produção |
| CADSUS | Acompanhamento das cargas do CADSUS oriundas dos módulos descentralizados. | Produção |
| DENGUE | Gestão e acompanhamento da evolução dos casos de DENGUE no Brasil. | Produção |
| CARTASUS | Gestão e acompanhamento das Cartas emitidas aos cidadãos que fizeram uso do SUS. | Produção |
| CONTROLE DAS AÇÕES DO TRABALHO | Controle incentivo financeiro para capacitação dos trabalhadores de vigilância sanitária dos Estados, Distrito Federal e Municípios. | Produção |
| BI DRAC-Apurações Especiais | Prover o DRAC de uma ferramenta construída com as técnicas de BI para apresentação das informações por meio de painéis de indicadores (*dashboards*) e de relatórios. | Produção |
| BOLSA ALIMENTAÇÃO | Acompanhamento do Sistema Bolsa Alimentação. | Produção |

Fonte: DATASUS B, 2015.

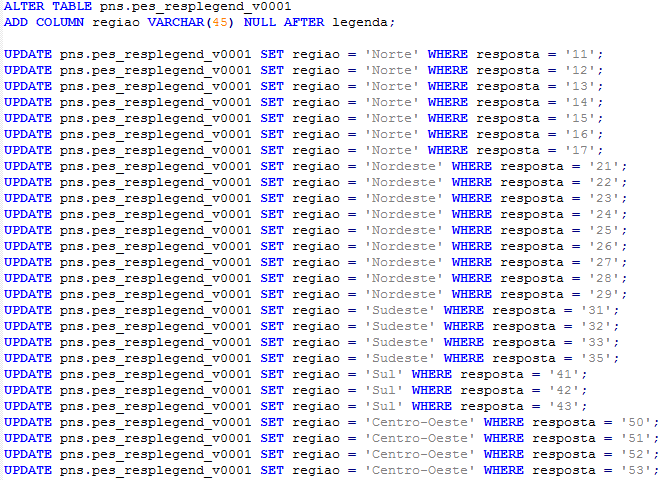
# APÊNDICE A – *SCRIPTS* PARA A CRIAÇÃO DO *DATA WAREHOUSE*

Figura - *Script*: 4-aux\_faixa\_etaria.sql



Fonte: Autor, 2016.

Figura - *Script*: 5-pes\_resplegend\_v0001.sql



Fonte: Autor, 2016.

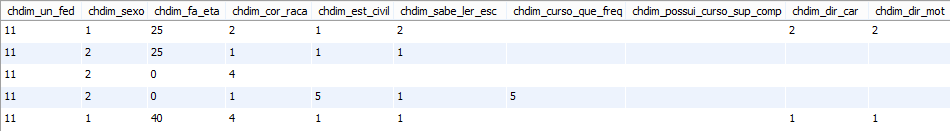
Figura - *Script*: 0-schema e usuario.sql



Fonte: Autor, 2016.

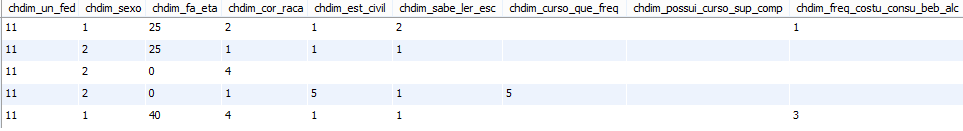
# APÊNDICE B – CONTEÚDO DAS *VIEWS* FATO E DIMENSÕES, LIMITADO A 5 REGISTROS

Figura - *Select* na *view* “fato\_acidentes”



Fonte: Autor, 2016.

Figura – *Select* na *view* “fato\_uso\_alcool”



Fonte: Autor, 2016.

Figura – *Select* na *view* “dim\_un\_fed”



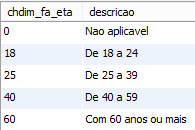
Fonte: Autor, 2016.

Figura – *Select* na *view* “dim\_sexo”



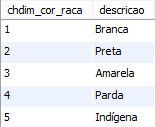
Fonte: Autor, 2016.

Figura – *Select* na *view* “dim\_fa\_eta”



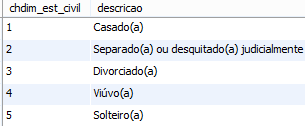
Fonte: Autor, 2016.

Figura – *Select* na *view* “dim\_cor\_raca”



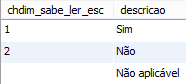
Fonte: Autor, 2016.

Figura – *Select* na *view* “dim\_est\_civil”



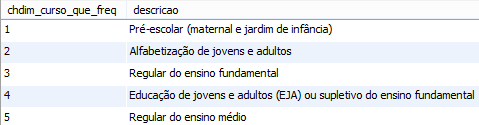
Fonte: Autor, 2016.

Figura – *Select* na *view* “dim\_sabe\_ler\_esc”



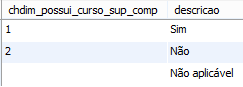
Fonte: Autor, 2016.

Figura – *Select* na *view* “dim\_curso\_que\_freq”



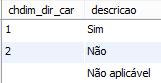
Fonte: Autor, 2016.

Figura *Select* na *view* “dim\_possui\_curso\_sup\_comp”



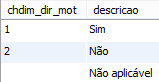
Fonte: Autor, 2016.

Figura – *Select* na *view* “dim\_dir\_car”



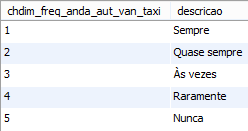
Fonte: Autor, 2016.

Figura – *Select* na *view* “dim\_dir\_mot”



Fonte: Autor, 2016.

Figura - *Select* na *view* “dim\_freq\_anda\_aut\_van\_taxi”



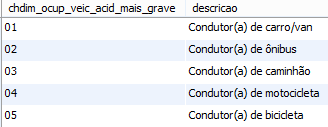
Fonte: Autor, 2016.

Figura - *Select* na *view* “dim\_envol\_acid\_c\_lesoes\_ult\_12m”



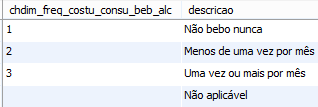
Fonte: Autor, 2016.

Figura - *Select* na *view* “dim\_ocup\_veic\_acid\_mais\_grave”



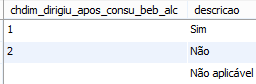
Fonte: Autor, 2016.

Figura - *Select* na *view* “dim\_freq\_costu\_consu\_beb\_alc”



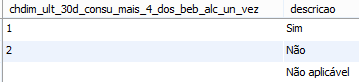
Fonte: Autor, 2016.

Figura - Select na view “dim\_dirigiu\_apos\_consu\_beb\_alc”



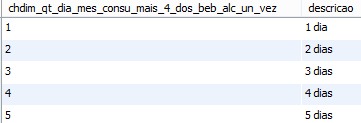
Fonte: Autor, 2016.

Figura - *Select* na *view* “dim\_ult\_30d\_consu\_mais\_4\_dos\_beb\_alc\_un\_vez”



Fonte: Autor, 2016.

Figura – *Select* na *view* “dim\_qt\_dia\_mes\_consu\_mais\_4\_dos\_beb\_alc\_un\_vez”



Fonte: Autor, 2016.