

4 Aplicação da Sistemática

Este capítulo descreve a aplicação da sistemática definida no Capítulo 3 utilizando dados reais de uma estatística pública e aplicando tecnologias avançadas fazendo o uso de XML e Serviços Web. O produto final dessa sistemática é um protótipo para prova-de-conceito de acordo com as diretrizes do grupo de interesse do e-Governo do W3C-Brasil.

Como o objetivo de criar um protótipo de uma estatística pública produzida por instituições de pesquisas governamentais brasileira fez-se a seleção com base nos seguintes critérios: Para começar, foi escolhida uma instituição de pesquisa governamental que disponibilizasse dados estatísticos na esfera federal na Web. Seguindo esse critério apareceram, várias instituições como Instituto Brasileiro de Geografia e Estatística (IBGE), Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), Departamento de informática do SUS (DATASUS) e Ministério do Trabalho. A fim de obter dados estatísticos com um grande escopo temático (por exemplo: saúde, educação, emprego) fez-se a opção pelos dados produzidos pelo IBGE. Um dos desafios era de aplicar essa sistemática em uma pesquisa em que a produção dela tivesse uma periodicidade regular para a análise de tendências em relação ao tempo, por isso, foi selecionada a Pesquisa Nacional por Amostra de Domicílios (PNAD).

4.1. Etapa 1 – Modelo Conceitual da Estatística

Após uma pesquisa no site oficial do IBGE e nos microdados foi possível obter informações para a modelagem conceitual da estatística.

A PNAD é um levantamento estatístico que integra o Programa Nacional de Pesquisas Contínuas por Amostra de Domicílios da Fundação IBGE, a PNAD vem sendo realizada desde 1967 com um duplo objetivo:

- Suprir a falta de informações sobre a população brasileira durante o período intercensitário;
- Estudar temas insuficientemente investigados ou não contemplados nos censos demográficos decenais realizados por aquela instituição.

Considerando a impossibilidade de investigar continuamente todos os temas de interesse, a PNAD foi estruturada para ter uma pesquisa básica, pesquisas suplementares e pesquisas especiais.

A pesquisa básica investiga, de forma contínua, os temas definidos como de maior importância para medir e acompanhar o nível socioeconômico da população: habitação e mão-de-obra, além de características demográficas e educacionais.

As pesquisas suplementares aprofundam os temas permanentes e investigam outros assuntos de interesse que se interliguem com os da pesquisa básica.

As pesquisas especiais abordam assuntos de maior complexidade, que exigem tratamento à parte da pesquisa básica, podendo até requerer um esquema de amostragem distinto.

Na aplicação dessa sistemática foi considerada apenas a pesquisa básica da PNAD para o ano referência 2008.

A pesquisa básica da PNAD abrange a população residente em domicílios particulares permanentes e em unidades de habitação em domicílios coletivos. A coleta de informações obedece a uma série de conceitos e definições operacionais, iguais ou assemelhados aos utilizados em várias outras pesquisas domiciliares, inclusive o Censo Demográfico, o que facilita sobremaneira a comparação dos indicadores produzidos por esse tipo de levantamento.

Após a coleta de dados referente a um período, o IBGE compila e disponibiliza na Web os microdados correspondentes ao período. Apenas alguns arquivos dos microdados foram utilizados ao longo da aplicação dessa sistemática e serão mencionados de acordo com o uso.

1. Identificar os elementos investigados pela estatística (fatos)

Elemento: Pessoa residente em um domicílio

2. Identificar a granularidade da pesquisa (grão)

Como a PNAD é realizada anualmente e tem por objetivo avaliar as condições de vida da população no ano, então a granularidade definida para o DW foi o “Ano de referência”

3. Identificar as dimensões a partir das variáveis de análise

Dentre os arquivos dos microdados disponibilizados no site do IBGE, utilizou-se o arquivo “dicionário de variáveis 2008.xls”. Este arquivo contém o dicionário de dados relacionando as variáveis de análise com a pergunta da pesquisa realizada pelo instituto. Para cada variável de análise, ou pergunta do questionário, foi gerada uma dimensão correspondente. Exemplos de dimensões para análise: “Sexo”, “Cor ou raça”, “Idade do morador na data de referência”.

4. Identificar as variáveis de medidas da pesquisa

A PNAD é uma pesquisa amostral, ou seja, a pesquisa é aplicada em uma parte representativa da população e através de cálculos estatísticos realizados pelo IBGE faz-se uma estimativa da população atribuindo pesos para cada questionário respondido. Como por exemplo, um questionário respondido por um morador do Rio de Janeiro, com 35 anos, cor branca, nível de escolaridade superior completo, renda superior a 5(cinco) salários mínimos tem um peso de 200 pessoas, ou seja, estima-se que 200 pessoas da população brasileira tem as mesmas características levantadas. Com isso, identificou-se a variável de medida da PNAD como o peso da pessoa.

Variável de medida: “V4729- Peso da pessoa”

5. Mapear as informações identificadas no esquema dimensional

Como são muitas dimensões com a mesma estrutura, então o esquema dimensional resultante está parcialmente mostrado na Figura 12.

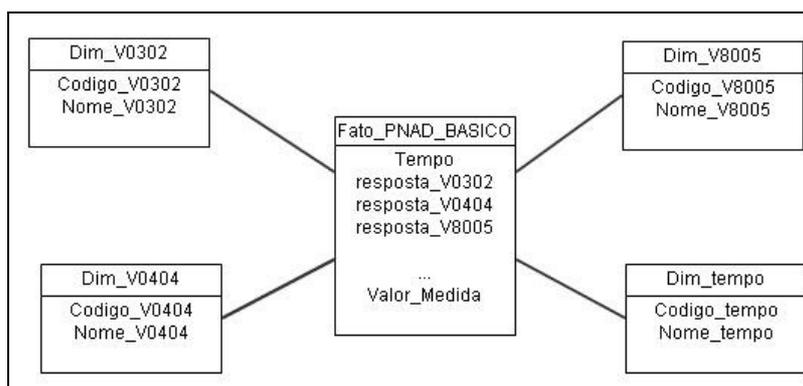


Figura 12 – Esquema dimensional resultante

4.2.

Etapa 2 – Especificação dos Requisitos OLAP

1. Agrupar as dimensões de acordo com a necessidade de hierarquia de análise. No caso da PNAD, foram definidos dois grupos para análise:

Dimensões sem transformação: esse grupo contém as variáveis estatísticas que não precisam de uma derivação de seus valores, ou seja, a análise deve ser exatamente igual ao dado coletado. Exemplos dessas variáveis são: V0302-Sexo, V0404-cor ou raça.

Dimensões com mais de 1 hierarquia: esse grupo contém as variáveis que devem ter apenas mais um nível de hierarquias. As derivações serão estabelecidas nos passos seguintes dessa etapa. Exemplos dessas variáveis são: “V8005-Idade” e “V1252-Rendimentos de aposentadoria”.

2. Criar os modelos de tabelas dimensões padrões para atender aos requisitos de análise.

A estrutura das dimensões sem alteração e com alteração esta na Figura 13.

3. Especificar as regras de derivação dos valores para as variáveis

Para as dimensões com mais de 1 hierarquia, as derivações das variáveis são:

- UF – criar as regiões (“Centro-Oeste”, “Nordeste”, “Norte”, “Sudeste”, “Sul”) de acordo com UF coletada.

- Variáveis do tipo moeda, devem ser transformadas em intervalos de salário mínimo, tais como: “Até ¼ salário mínimo”, “Mais de ¼ até ½ salário”, “Mais de ½ até 1 salário”, “Mais de 1 até 2 salários”, “Mais de 2 até 3 salários”, “Mais de 3 até 5 salários”, “Mais de 5 salários”.
- A variável “V8005-Idade do morador da data de referência” deve ser derivada para as faixas de idade. No caso deste protótipo foi utilizado intervalos de 5 anos, tais como: “0 a 4”, “5 a 9”, “10 a 14”,....., “65 a 69” e “70 ou mais”

4. Classificar as variáveis de medida quanto a aditividade

A variável de medida identificada na etapa 1 foi a “V4729- Peso pessoa” e ela é semi-aditiva em relação a dimensão tempo “V0101-Ano referência”, porque, não faz sentido somar a população residente em anos distintos.

Dessa forma, a etapa se encerra com a definição das estruturas das tabelas dimensões conforme a Figura 13 abaixo e a derivações necessárias para a dimensões foram acrescentadas mais uma hierarquia.

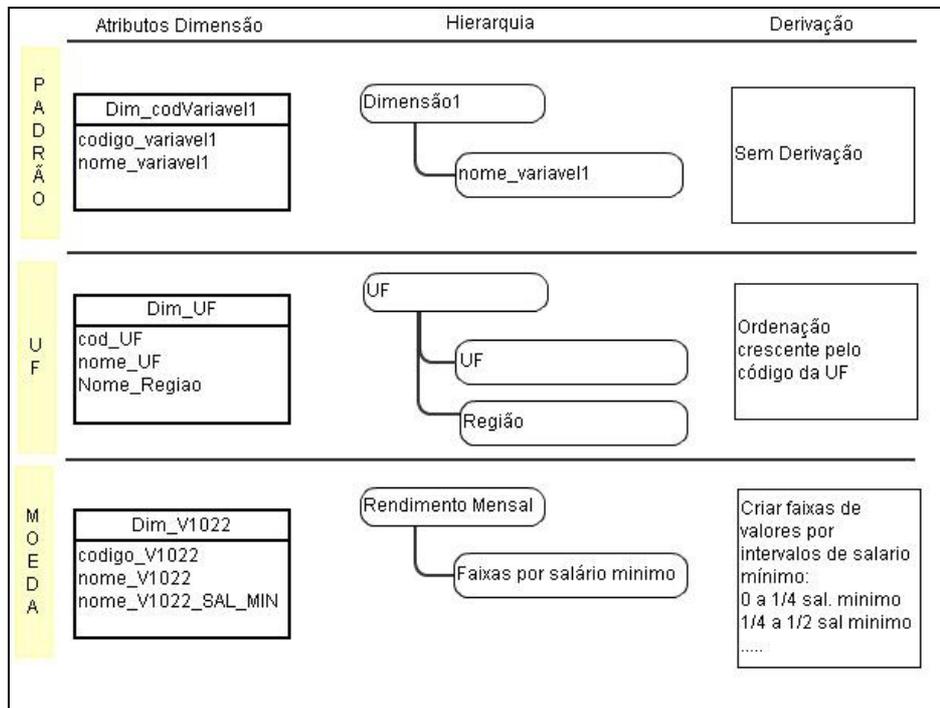


Figura 13 - Definição das estruturas das tabelas dimensões

4.3.

Etapa 3- Construção do Esquema Lógico

1. Adequar a nomenclatura

A nomenclatura das tabelas de dimensões deve ter o prefixo “DIM_” e o código da variável (i.e. DIM_V0302) e da tabela fato “FATO_” e a sigla da pesquisa, no caso, FATO_PNAD_BASICA.

2. Criar as chaves primárias

As tabelas dimensões devem ter um identificador único (*primary key*) e a tabela FATO_PNAD_BASICA deve ter um identificador numérico seqüencial (*surrogate key*) para utilização interna e as chaves estrangeiras (*foreign key*) em todos os atributos, exceto nos atributos de medidas.

O esquema lógico completo da PNAD contém mais de 200 tabelas de dimensões com a mesma estrutura, por isso, o esquema foi parcialmente representado na Figura 14.

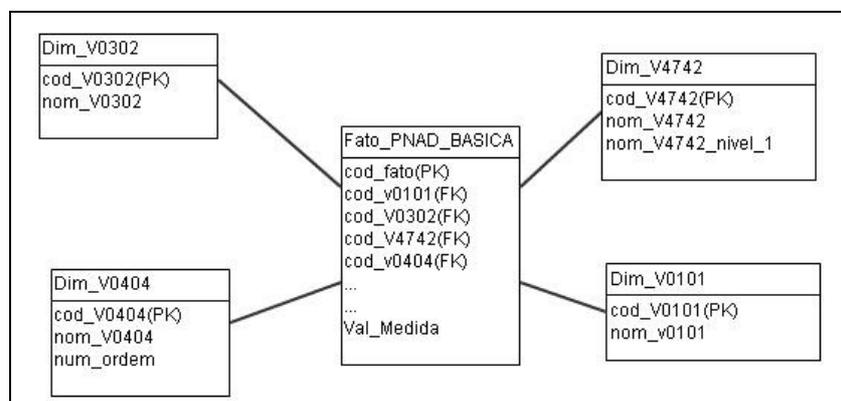


Figura 14 - Esquema lógico Dimensional da PNAD Básica

4.4.

Etapa 4 - Metadados

O foco dessa etapa é construir o modelo dos metadados semânticos para serem levados até o cubo, por isso foi definida uma estrutura mínima para armazenar a contextualização de construção de cada variável e suas categorias

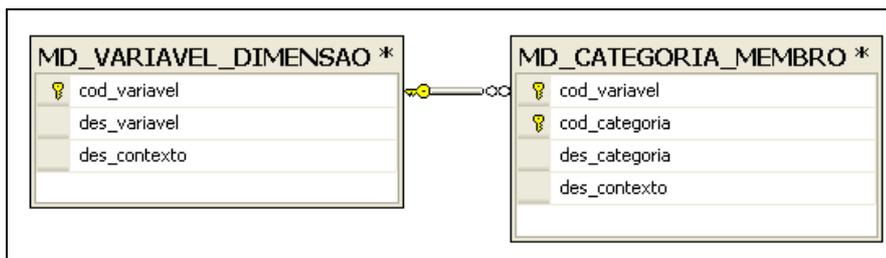


Figura 15 - Modelo Físico metadados semânticos

1. Descrição das tabelas e seus atributos:

Tabela	
MD_VARIAVEL_DIMENSAO	Armazena as variáveis das PNAD que se tornarão as dimensões
Atributos	
cod_variavel	Chave primária única, sendo um o código da variável
des_variavel	Descrição da variável (a pergunta realizada no questionário)
des_contexto	Notas metodologicas referente a variável

Tabela	
MD_CATEGORIA_MEMBRO	Armazena as categorias das variáveis que vão se tornar os membros das dimensões..
Atributos	
cod_variavel	Chave composta, sendo um o código da variável
cod_categoria	Chave composta, sendo os categorias das variáveis, (o conjunto de respostas de uma pergunta)
des_categoria	Código da variável observada ou medida.
des_contexto	Notas metodológicas referentes a categoria

Essas tabelas foram populadas a partir dos arquivos “Dicionário de variáveis de pessoas-2008.xls” e “Notas Metodológicas - Pesquisa Básica, Especial e Suplementares PNAD 2008.doc” disponibilizados nos microdados da PNAD.

4.5.

Etapa 5- Definição da Arquitetura

Para realizar as atividades desenvolvidas nessa sistemática foram utilizados softwares licenciados porque a autora tem maior habilidade neles, porém toda a solução pode ser desenvolvida utilizando softwares *open source* (desde o sistema operacional, SGBD e servidor OLAP até às ferramentas Web disponíveis para os clientes).

O requisito crítico na arquitetura proposta é o servidor OLAP, deve-se utilizar um servidor OLAP que execute queries MDX e que funcione como um XMLA *provider*, obedecendo assim às especificações do XMLA *Council*. Exemplos de atuais softwares que atendem a esses critérios são: Microsoft Analysis Services, SAP BW e o Mondrian, sendo esse último um *open source* gratuito. O Mondrian também é parte integrante de um suíte de soluções de BI denominada Pentaho BI suite. Essa suíte disponibiliza ferramentas de qualidade igual ou superior as soluções comerciais conhecidas, englobando funcionalidades desde módulos de geração de relatórios a ferramentas de Mineração de Dados (Arriaga, 2008).

1. Identificar os arquivos fontes

Dentre os arquivos dos microdados, utilizou-se o arquivo “PES2008.txt” que contém os dados coletados na pesquisa e o arquivo “dicionário de variáveis de pessoa 2008.xls” que contém os metadados da pesquisa. Sem este último, não é possível realizar a leitura dos dados obtidos na PNAD. Já, o arquivo “Notas Metodológicas - Pesquisa Básica, Especial e Suplementares PNAD 2008.doc” contém a conceitos necessários para a contextualização das dimensões e seus membros.

2. Definir as ferramentas para o desenvolvimento do ETL

Tanto para o processo de ETL, quanto para a construção do modelo físico do DW, decidiu-se utilizar *scripts* SQL¹², e para a construção do Cubo, utilizou-se a linguagem XMLA. Esta solução permite uma maior flexibilidade para definir

¹² Structured Query Language

o Sistema gerenciador de Banco de dados (SGBD), mas tem um custo de manutenção maior do que utilizar uma ferramenta de ETL mais poderosa para realizar o ETL..

Como os *scripts* SQL manipulam apenas os dados que estejam dentro de um SGBD, então no Microsoft SQL Server 2005 foi criado um banco de dados intermediário denominado “ODS_ESTATISTICA” com as estruturas das tabelas semelhantes aos arquivos fontes e os arquivos foram migrados para esse banco utilizando as ferramentas desse SGBD.

3. Definir o repositório para os Metadados e Data Warehouse

Com o SGBD, Microsoft SQL Server 2005, foi instanciado o banco “DW_ESTATISTICA” para armazenar os esquemas estrelas e as tabelas de metadados.

4. Definir o servidor OLAP

Com o servidor OLAP, Microsoft SQL Server 2005 Analysis Services (SSAS), foi instanciado o banco multidimensional “OLAP_ESTATISTICA”.

5. Configurar e testar todos os elementos da arquitetura

O diagrama da arquitetura está representado na Figura 16.

Figura 16 - Diagrama da arquitetura na aplicação da sistemática PNAD Básica

4.6.

Etapa 6 - Projeto de ETL e desenvolvimento

A coleta de dados da PNAD é realizada anualmente oferecendo a visualização dos dados ao final de um período. Por isso, a carga da tabela fato deve ser “instantâneo periódico”. A lógica do processo de carga deve garantir a integridade referencial da tabela fato com as dimensões para qualquer período de tempo.

Preparação das fontes

1. Acessar e identificar as fontes de dados

O arquivo de dados “Pes2008.txt” tem 399.890 linhas e cada linha representa as respostas dos questionários fechado de uma pessoa moradora de unidade domiciliar. Nesse arquivo, os dados estão dispostos em colunas lado a lado sem nenhum tipo de separação. Um pequeno trecho do arquivo é apresentado na Figura 17.

```
20071100014701807411061996011432812062...3426 121 3262623 122011 7 ...
20071100014701808228061996011432812062...3426 121 3262623 122011 5 ...
20071100014701809213112002004432812062...3426 121 5 2623 32209 ...
...
```

Figura 17 – Trecho do arquivo “Pes2008.txt”

O arquivo de metadados “Dicionário de variáveis de pessoas-2008.xls”, cujo formato é representado na Tabela 17, contém as variáveis de análise e medida da pesquisa, o conjunto de valores possíveis de cada variável e a localização e dimensão física do dado no arquivo “Pes2008.txt”.

As colunas POSIÇÃO INICIAL e TAMANHO determinam a localização e dimensão física do dado no arquivo “Pes2008.txt”. A característica pesquisada é identificada através de três campos: CÓDIGO DE VARIÁVEL, QUESITO N°(ordem de coleta da variável), QUESITO DESCRIÇÃO (a pergunta realizada na coleta). Os campos CATEGORIA TIPO e CATEGORIA DESCRIÇÃO representam o código e a descrição do conjunto de dados observados em uma característica.

Posição Inicial	Tamanho	Código de variável	Quesito		Categorias	
			N°	Descrição	Tipo	Descrição
33	1	V0404	4	Cor ou raça	2	Branca
					4	Preta
					6	Amarela
					8	Parda
					0	Indígena
					9	Sem declaração

Tabela 2 - Formato do arquivo de metadados da PNAD

2. Padronizar os arquivos de metadados da pesquisa

Com o objetivo de preparar os arquivos fontes para iniciar a extração, o arquivo “Dicionário de variáveis de pessoas - 2008.xls” passou por um processo de padronização. As linhas das variáveis da pesquisa suplementar foram excluídas. As células que estavam mescladas foram desmescladas e inserido o valor correspondente. Na coluna TIPO_CATEGORIA para as células vazias foi inserido uma sequência de caracteres “ZZZ” de tamanho correspondente ao tamanho definido na variável. O formato final pode ser verificado na Tabela 3.

Posição inicial	tamanho	código variavel	Quesito	Código categoria	Categoria
33	1	V0404	Cor ou raça	2	Branca
33	1	V0404	Cor ou raça	4	Preta
33	1	V0404	Cor ou raça	6	Amarela
33	1	V0404	Cor ou raça	8	Parda
33	1	V0404	Cor ou raça	0	Indígena
33	1	V0404	Cor ou raça	9	Sem declaração
34	1	V0405	Tem mãe viva	1	Sim
34	1	V0405	Tem mãe viva	3	Não
34	1	V0405	Tem mãe viva	5	Não sabe
35	1	V0406	Mãe mora no domicílio	2	Sim
35	1	V0406	Mãe mora no domicílio	4	Não
35	1	V0406	Mãe mora no domicílio	Z	Não aplicável

Tabela 3 - Formato do arquivo de metadados após padronização

3. Construir o modelo físico dos metadados

No banco de dados intermediário (ODS_ESTATISTICA) foram criadas duas tabelas com as estruturas das tabelas semelhantes aos arquivos fontes. Dessa forma, os arquivos fontes da PNAD foram importados para o ODS_ESTATISTICA respeitando suas formatações originais. A tabela `dicionario_de_variaveis_pessoa` recebeu os dados do arquivo “Dicionário de variáveis de pessoas-2008.xls” e a tabela `PES2008` recebeu os dados do arquivo “Pes2008.txt”. Nesta tabela foi criada uma chave primária única, sendo atribuído um número inteiro seqüencial.

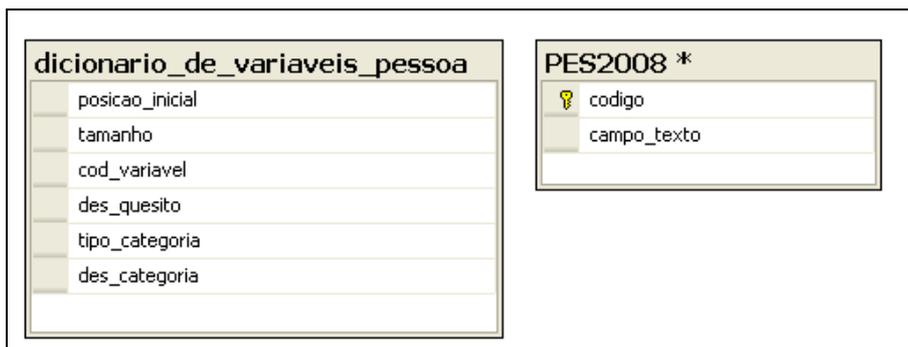


Figura 18 – Modelo físico das fontes de dados importados para o ODS

Descrição das tabelas e seus atributos:

Tabela	
Pes2008	Armazena os dados coletados na PNAD contidos nas linhas do arquivo Pes2008.txt
Atributos	
Código	Chave primária única, sendo um número inteiro atribuído seqüencialmente
Campo_texto	Linha do arquivo texto com 809 caracteres

Tabela 4 - Tabela pess2008 correspondente aos dados da PNAD

Tabela	
Dicionário_de_variaveis_pessoa	Armazena os metadados da pesquisa necessários para a leitura do arquivo Pes2007.txt e os metadados para a geração do modelo físico do Data Mart.
Atributos	
posicao_inicial	Indica a posição, na cadeia de caracteres do arquivo texto, que inicia a representação do dado referente à característica observada ou medida
Tamanho	Indica o tamanho da cadeia de caracteres do dado
cod_variavel	Código da variável observada ou medida.
num_quesito	Indica a ordem em que a variável foi coletada
des_quesito	A pergunta realizada para a coleta
tipo_categoria	Código do dado
des_categoria	Descrição do dado

Tabela 5 - Dicionário de variáveis correspondente aos metadados da PNAD

4. Importar os microdados da pesquisa para os metadados do DW

Com os dados no ODS_ESTATISTICA, populou-se as tabelas dos metadados e iniciou o desenvolvimento e execução de vários scripts SQL. Em linhas gerais, cada script realiza uma consulta no ODS_ESTATISTICA e o resultado é um conjunto de instruções SQL/XMLA que permitem criar/alterar objetos ou manipular/inserir dados nos bancos de dados. Durante a execução dos scripts, sempre que alguma anomalia for identificada, a ação deve ser abortar completamente o processo exigindo uma intervenção manual para identificação e correção do problema.

Para a construção dos metadados do DW foram definidas as tabelas dos metadados, conforme a Figura 18. O arquivo “Notas Metodológicas - Pesquisa Básica, Especial e Suplementares PNAD 2008.doc” contém a contextualização de cada variável ou categoria definida para a coleta. Assim, para a carga da tabela, os metadados foram manualmente extraídos e inserido nas tabelas correspondentes. Esse procedimento foi realizado para algumas variáveis e categorias da pesquisa apenas para a demonstração desse trabalho.

Construção do modelo físico do DW

Os scripts em SQL lêem do ODS_ESTATISTICA as variáveis com suas estruturas e cria uma tabela temporária contendo os metadados, em seguida, é realizada uma segunda *query* gerando as instruções DDL¹³ para criar os objetos no DW_ESTATISTICA. Todos os scripts utilizados no ETL estão no Anexo 1.

5. Criar a estrutura básica das tabelas dimensões

Para construir as tabelas dimensões, primeiramente, foi executado o *script* “Cria tabelas dimensoes.sql”, que esta no Anexo 1. Vale observar, que todas as tabelas dimensões são criadas inicialmente com a estrutura básica e após a carga que elas devem ser alteradas.

¹³ DDL - Data Definition Language

6. Criar tabela fato

A construção da tabela fato também foi realizada através de um *script* SQL denominado “Cria tabela fato.sql” que esta no Anexo 1.

Limpeza e transformação na área temporária

Os scripts em SQL lêem do ODS_ESTATISTICA as variáveis criando uma tabela temporária contendo os metadados e uma segunda *query* monta as instruções DML¹⁴ para inserir os dados nas tabelas correspondentes no DW_ESTATISTICA. Todos os scripts utilizados na aplicação dessa sistemática estão no Anexo 1.

7. Realizar a carga da tabela fato

Executou-se o *script* “Carga_tabela_fato.sql”.

8. Realizar a carga das tabelas dimensões

“Carga_dimensoes.sql”. Realiza a carga das tabelas dimensões a partir dos metadados e inclusive cria categorias quando elas foram declaradas como intervalos. Por exemplo, a variável “V9038 - Número de dias na semana que prestava serviço domestico” onde as categorias disponíveis são “1 a 7” e não os valores individuais “1”, “2”, “3”, ..., “7”.

9. Definir cenários para validar a integridade dos dados

Para validação dos dados do período, foram definidos os seguintes cenários implementados no script “Valida_dados.sql”.

Teste de validação de uma única coluna:

- Tabela FATO: Cada campo com valor nulo, foi substituído pela seqüência de caracteres “ZZZZZ” de acordo com o tamanho de cada campo.
- Tabelas DIMENSÃO: Verificar em cada tabela dimensão se o campo que será uma PK tem alguma ocorrência com valor nulo.

¹⁴ DML - Data Manipulation Language

Teste pelo cruzamento entre colunas:

- Integridade Referencial: Insere nas tabelas temporárias de dimensões os códigos de categorias que existem na tabela temporária fato e não existem nas tabelas dimensões, utilizando as descrições da primeira ocorrência da tabela de dimensões.

Após a integridade referencial estar validada, integridade referencial foi inserida nas tabelas e em seguida as tabelas de dimensões foram alteradas de acordo com as necessidades de hierarquias.

10. Incluir as PK nas tabelas dimensões e fato

“Inclui_PK_FK.sql”. A inclusão das PK e FK foi executada após a carga e validação dos dados nas tabelas fato e dimensões.

11. Alterar as tabelas dimensões incluindo as hierarquias de análise necessárias para cada grupo de dimensão, executa as regras de derivação e popula as tabelas

“Altera_e_popula_tabela_dimensoes.sql”.

Construção do Cubo

Para a construção, primeiramente foi necessário criar as dimensões e em seguida criou-se o cubo com as dimensões existentes. Para criar as dimensões foram desenvolvidos e executados scripts que lêem do ODS_ESTATISTICA e dos metadados do DW as variáveis e seus contextos gerando como resultado um conjunto de instruções em XMLA encapsulado no método execute do XMLA. O resultado foi manualmente selecionado e executado em uma sessão do XMLA do servidor OLAP. Para isso, o *script* SQL executado foi o “Cria_dimensoes_cubos.sql” .

Para construir o cubo foi utilizado editor de cubos “Microsoft SQL Server BI Development Studio”. Nesta etapa, foi necessário estabelecer que a medida, no caso da PNAD, deve ser semi-aditivas em relação a dimensão tempo. Em seguida, o cubo foi processado na mesma ferramenta.

17. Cenários para validar as informações analíticas geradas

Para validar a qualidade da informação gerada, após o processamento do cubo, cenário idênticos as publicações disponibilizadas na pagina IBGE foram criados os os valores comparados.

Os cenários foram criados no cubo para conferência com a publicação do IBGE “Pesquisa Nacional por Amostra de Domicílio”, volume 29, 2008, Brasil.

IBGE		CUBO “demo”																							
Total	57 347	<table border="1"> <thead> <tr> <th>Estimativa População</th> <th colspan="2">V0302-Sexo</th> <th>Total geral</th> </tr> <tr> <th>V6002-Rede de ensino</th> <th>Feminino</th> <th>Masculino</th> <th>Total geral</th> </tr> </thead> <tbody> <tr> <td>Particular</td> <td>6.508.922</td> <td>5.802.702</td> <td>12.311.624</td> </tr> <tr> <td>Pública</td> <td>22.456.199</td> <td>22.579.177</td> <td>45.035.376</td> </tr> <tr> <td>Total geral</td> <td>28.965.121</td> <td>28.381.879</td> <td>57.347.000</td> </tr> </tbody> </table>				Estimativa População	V0302-Sexo		Total geral	V6002-Rede de ensino	Feminino	Masculino	Total geral	Particular	6.508.922	5.802.702	12.311.624	Pública	22.456.199	22.579.177	45.035.376	Total geral	28.965.121	28.381.879	57.347.000
Estimativa População	V0302-Sexo		Total geral																						
V6002-Rede de ensino	Feminino	Masculino	Total geral																						
Particular	6.508.922	5.802.702	12.311.624																						
Pública	22.456.199	22.579.177	45.035.376																						
Total geral	28.965.121	28.381.879	57.347.000																						
Rede de ensino																									
Pública	45 035																								
Particular	12 312																								

Figura 19 - Pessoas que freqüentavam creche ou escola, segundo a rede de ensino - Brasil – 2008

IBGE				CUBO “demo”																																																																																																																										
<table border="1"> <thead> <tr> <th rowspan="2">Grupos de idade</th> <th colspan="3"></th> </tr> <tr> <th>Total</th> <th>Homens</th> <th>Mulheres</th> </tr> </thead> <tbody> <tr> <td>Total</td> <td>189 953</td> <td>92 433</td> <td>97 520</td> </tr> <tr> <td>0 a 4 anos</td> <td>13 622</td> <td>6 910</td> <td>6 712</td> </tr> <tr> <td>Menos de 1 ano</td> <td>2 553</td> <td>1 299</td> <td>1 254</td> </tr> <tr> <td>1 a 4 anos</td> <td>11 069</td> <td>5 611</td> <td>5 458</td> </tr> <tr> <td>5 a 9 anos</td> <td>15 770</td> <td>8 028</td> <td>7 743</td> </tr> <tr> <td>10 a 14 anos</td> <td>17 562</td> <td>8 958</td> <td>8 605</td> </tr> <tr> <td>15 a 19 anos</td> <td>16 970</td> <td>8 713</td> <td>8 257</td> </tr> <tr> <td>15 a 17 anos</td> <td>10 290</td> <td>5 300</td> <td>4 990</td> </tr> <tr> <td>18 ou 19 anos</td> <td>6 681</td> <td>3 413</td> <td>3 267</td> </tr> <tr> <td>20 a 24 anos</td> <td>16 561</td> <td>8 355</td> <td>8 207</td> </tr> </tbody> </table>				Grupos de idade				Total	Homens	Mulheres	Total	189 953	92 433	97 520	0 a 4 anos	13 622	6 910	6 712	Menos de 1 ano	2 553	1 299	1 254	1 a 4 anos	11 069	5 611	5 458	5 a 9 anos	15 770	8 028	7 743	10 a 14 anos	17 562	8 958	8 605	15 a 19 anos	16 970	8 713	8 257	15 a 17 anos	10 290	5 300	4 990	18 ou 19 anos	6 681	3 413	3 267	20 a 24 anos	16 561	8 355	8 207	<table border="1"> <thead> <tr> <th>Estimativa População</th> <th colspan="2">V0302-Sexo</th> <th>Total geral</th> </tr> <tr> <th>V8005-Faixa-Idade</th> <th>Feminino</th> <th>Masculino</th> <th>Total geral</th> </tr> </thead> <tbody> <tr> <td>00 a 04 anos</td> <td>6.711.789</td> <td>6.909.942</td> <td>13.621.731</td> </tr> <tr> <td>05 a 09 anos</td> <td>7.742.669</td> <td>8.027.584</td> <td>15.770.253</td> </tr> <tr> <td>10 a 14 anos</td> <td>8.604.700</td> <td>8.957.501</td> <td>17.562.201</td> </tr> <tr> <td>15 a 19 anos</td> <td>8.257.331</td> <td>8.712.974</td> <td>16.970.305</td> </tr> <tr> <td>20 a 24 anos</td> <td>8.206.724</td> <td>8.354.690</td> <td>16.561.414</td> </tr> <tr> <td>25 a 29 anos</td> <td>8.318.683</td> <td>7.920.722</td> <td>16.239.405</td> </tr> <tr> <td>30 a 34 anos</td> <td>7.548.788</td> <td>7.052.706</td> <td>14.601.494</td> </tr> <tr> <td>35 a 39 anos</td> <td>7.050.534</td> <td>6.448.093</td> <td>13.498.627</td> </tr> <tr> <td>40 a 44 anos</td> <td>7.042.994</td> <td>6.437.914</td> <td>13.480.908</td> </tr> <tr> <td>45 a 49 anos</td> <td>6.248.550</td> <td>5.647.661</td> <td>11.896.211</td> </tr> <tr> <td>50 a 54 anos</td> <td>5.590.309</td> <td>4.895.129</td> <td>10.485.438</td> </tr> <tr> <td>55 a 59 anos</td> <td>4.372.625</td> <td>3.853.099</td> <td>8.225.724</td> </tr> <tr> <td>60 a 64 anos</td> <td>3.546.285</td> <td>2.960.463</td> <td>6.506.748</td> </tr> <tr> <td>65 a 69 anos</td> <td>2.851.289</td> <td>2.307.371</td> <td>5.158.660</td> </tr> <tr> <td>70 ou mais</td> <td>5.426.968</td> <td>3.946.708</td> <td>9.373.676</td> </tr> <tr> <td>Total geral</td> <td>97.520.238</td> <td>92.432.557</td> <td>189.952.795</td> </tr> </tbody> </table>				Estimativa População	V0302-Sexo		Total geral	V8005-Faixa-Idade	Feminino	Masculino	Total geral	00 a 04 anos	6.711.789	6.909.942	13.621.731	05 a 09 anos	7.742.669	8.027.584	15.770.253	10 a 14 anos	8.604.700	8.957.501	17.562.201	15 a 19 anos	8.257.331	8.712.974	16.970.305	20 a 24 anos	8.206.724	8.354.690	16.561.414	25 a 29 anos	8.318.683	7.920.722	16.239.405	30 a 34 anos	7.548.788	7.052.706	14.601.494	35 a 39 anos	7.050.534	6.448.093	13.498.627	40 a 44 anos	7.042.994	6.437.914	13.480.908	45 a 49 anos	6.248.550	5.647.661	11.896.211	50 a 54 anos	5.590.309	4.895.129	10.485.438	55 a 59 anos	4.372.625	3.853.099	8.225.724	60 a 64 anos	3.546.285	2.960.463	6.506.748	65 a 69 anos	2.851.289	2.307.371	5.158.660	70 ou mais	5.426.968	3.946.708	9.373.676	Total geral	97.520.238	92.432.557	189.952.795
Grupos de idade																																																																																																																														
	Total	Homens	Mulheres																																																																																																																											
Total	189 953	92 433	97 520																																																																																																																											
0 a 4 anos	13 622	6 910	6 712																																																																																																																											
Menos de 1 ano	2 553	1 299	1 254																																																																																																																											
1 a 4 anos	11 069	5 611	5 458																																																																																																																											
5 a 9 anos	15 770	8 028	7 743																																																																																																																											
10 a 14 anos	17 562	8 958	8 605																																																																																																																											
15 a 19 anos	16 970	8 713	8 257																																																																																																																											
15 a 17 anos	10 290	5 300	4 990																																																																																																																											
18 ou 19 anos	6 681	3 413	3 267																																																																																																																											
20 a 24 anos	16 561	8 355	8 207																																																																																																																											
Estimativa População	V0302-Sexo		Total geral																																																																																																																											
V8005-Faixa-Idade	Feminino	Masculino	Total geral																																																																																																																											
00 a 04 anos	6.711.789	6.909.942	13.621.731																																																																																																																											
05 a 09 anos	7.742.669	8.027.584	15.770.253																																																																																																																											
10 a 14 anos	8.604.700	8.957.501	17.562.201																																																																																																																											
15 a 19 anos	8.257.331	8.712.974	16.970.305																																																																																																																											
20 a 24 anos	8.206.724	8.354.690	16.561.414																																																																																																																											
25 a 29 anos	8.318.683	7.920.722	16.239.405																																																																																																																											
30 a 34 anos	7.548.788	7.052.706	14.601.494																																																																																																																											
35 a 39 anos	7.050.534	6.448.093	13.498.627																																																																																																																											
40 a 44 anos	7.042.994	6.437.914	13.480.908																																																																																																																											
45 a 49 anos	6.248.550	5.647.661	11.896.211																																																																																																																											
50 a 54 anos	5.590.309	4.895.129	10.485.438																																																																																																																											
55 a 59 anos	4.372.625	3.853.099	8.225.724																																																																																																																											
60 a 64 anos	3.546.285	2.960.463	6.506.748																																																																																																																											
65 a 69 anos	2.851.289	2.307.371	5.158.660																																																																																																																											
70 ou mais	5.426.968	3.946.708	9.373.676																																																																																																																											
Total geral	97.520.238	92.432.557	189.952.795																																																																																																																											

Figura 20 - População residente, por sexo, segundo os grupos de idade - Brasil – 2008

4.7. Etapa 7 – Acesso via Web

Depois que o cubo foi processado no servidor OLAP, foi necessário configurar o servidor para permitir que o usuário final acesse os cubos através do protocolo HTTP. Tais configurações seguem as normas disponíveis no site da Microsoft.

Concluídas as configurações do servidor o cubo tornou-se disponível para ser acessado por qualquer aplicação, seja web ou desktop. que tenha suporte ao XMLA.

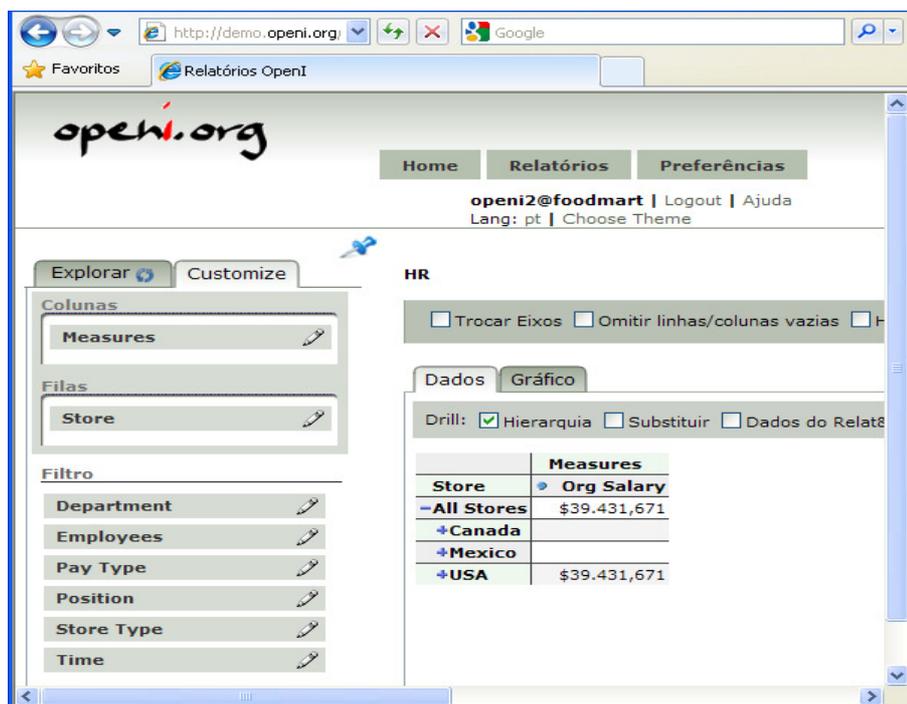
Vale ressaltar que não é possível acessar o cubo apenas por um web browser e sim através de uma aplicação. Como exemplos, temos:

I-Uso do Excel 2007(desktop)

Essa opção fornece uma facilidade tradicional para consumo das informações OLAP e permite a conexão XMLA.

II- Uso de aplicação Web (“open source”)

Por exemplo, as aplicações OpenI apresentada na Figura 21 e a Palo Web Cliente na Figura 22, também permitem o acesso via XMLA.



The screenshot shows the OpenI web application interface. The browser address bar displays 'http://demo.openi.org/'. The page header includes the 'openi.org' logo and navigation tabs for 'Home', 'Relatórios', and 'Preferências'. The user is logged in as 'openi2@foodmart' with options for 'Logout', 'Ajuda', 'Lang: pt', and 'Choose Theme'. The main content area is titled 'HR' and features a 'Customize' panel on the left with sections for 'Colunas' (Measures), 'Filas' (Store), and 'Filtro' (Department, Employees, Pay Type, Position, Store Type, Time). The main data area shows a table with columns 'Store' and 'Measures' (Org Salary). The data is presented in a hierarchical drill-down format.

Store	Measures
All Stores	\$39.431,671
+Canada	
+Mexico	
+USA	\$39.431,671

Figura 21 - Aplicação Web “open source” OpenI para acesso OLAP

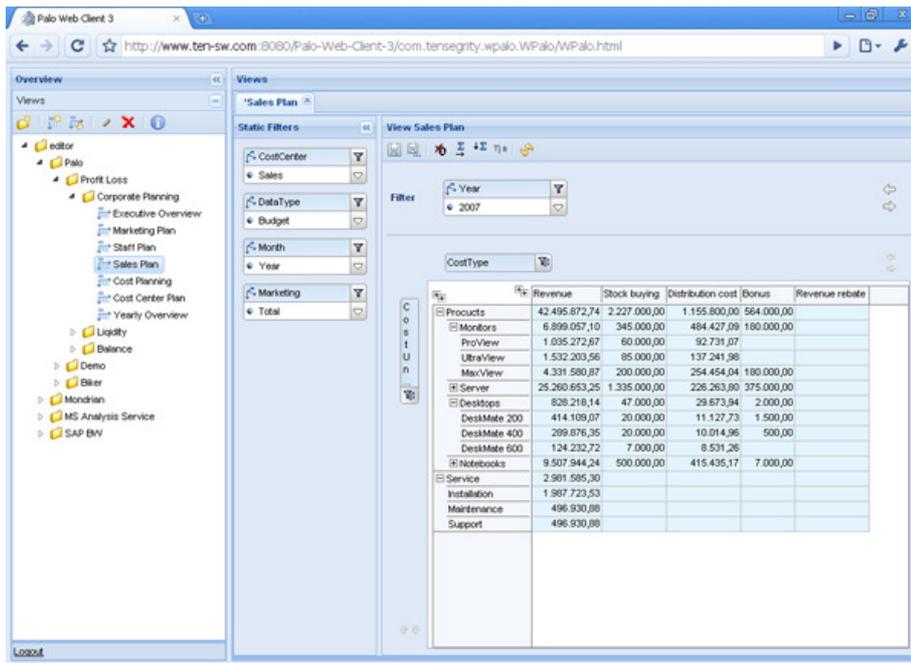


Figura 22 - Aplicação “open source” Palo Web Cliente para acesso OLAP

No próximo capítulo do trabalho será demonstrada a usabilidade das informações estatísticas a partir da versão “demo” gerada pela aplicação dessa sistemática.