

Flávia Andrade Rabelo

**Uma Sistemática Baseada em Data Warehousing
para Apoio ao Governo Eletrônico**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Informática da PUC-Rio.

Orientador: Prof. Rubens Nascimento Melo

Rio de Janeiro
Agosto de 2010

Flávia Andrade Rabelo

**Uma Sistemática Baseada em
Data Warehousing para Apoio
ao Governo Eletrônico**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico e Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Rubens Nascimento Melo

Orientador

Departamento de Informática – PUC-Rio

Prof. Julio Cesar Sampaio do Prado Leite

Departamento de Informática – PUC-Rio

Prof. Sean Wolfgang Matsui Siqueira

Departamento de Informática Aplicada – UNIRIO

Prof. José Eugenio Leal

Coordenador Setorial do Centro

Técnico Científico – PUC-Rio

Rio de Janeiro, 30 de agosto de 2010

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador

Flávia Andrade Rabelo

Graduada em Tecnólogo em Processamento de Dados em 1998 pela Universidade Católica de Brasília - UCB. Sua área de interesse acadêmico é Banco de Dados

Ficha Catalográfica

Rabelo, Flávia Andrade

Uma sistemática baseada em data warehousing para apoio ao governo eletrônico / Flávia Andrade Rabelo ; orientador: Rubens Nascimento Melo. – 2010.

123 f.: il. (color.) ; 30 cm

Dissertação (Mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2010.

Inclui bibliografia

1. Informática – Teses. 2. Data warehousing. 3. Governo eletrônico. 4. SaaS. 5. PNAD. I. Melo, Rubens Nascimento. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

André e Arthur, com muito amor

Agradecimentos

Agradeço à todos que contribuíram direta ou indiretamente para que eu pudesse concluir essa etapa. Primeiramente, ao meu orientador professor Rubens, por sempre ter acreditado no meu potencial e ter me guiado até aqui. Muito obrigada!

Aos professores Julio Cesar Sampaio do Prado Leite, Sean Wolfgang Matsui Siqueira por aceitarem participar dessa banca e contribuírem para a melhoria dessa pesquisa.

Aos amigos Jairo, Elaine e Tatiana pelo auxílio, orientação e suporte.

Aos meus amigos do Tribunal Regional Federal, que me incentivaram e compreenderam as minhas faltas e atrasos, em especial, ao meu querido e admirável chefe Roberto.

Aos meus amigos e familiares, próximos e distantes, que sempre me deram uma palavra de carinho me ensinando a superar as dificuldades.

À CAPES, pelo apoio financeiro, sem o qual este trabalho não poderia ser viabilizado.

À Deus, por tudo isso.

Resumo

Rabelo, Flávia Andrade; Melo, Rubens Nascimento. **Uma Sistemática baseada em Data Warehousing para apoio ao Governo Eletrônico**. Rio de Janeiro, 2010. 123 p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro

Com o objetivo de se adaptarem à nova forma de relacionamento entre os governos e a sociedade, os governos se deparam com a crescente necessidade de oferecer facilidade no acesso às informações públicas a um público cada vez mais diversificado. Com vistas a contribuir com esse propósito, o objetivo principal desta dissertação consiste no desenvolvimento de uma sistemática baseada em Data Warehousing para apoio ao governo eletrônico. Dessa forma busca-se oferecer um serviço de acesso analítico às informações para a sociedade via web e com baixo custo. Como prova de conceito da sistemática proposta, são utilizados dados reais de uma estatística pública (PNAD) disponibilizada na web. A aplicação da sistemática a esses dados gera um serviço de acesso analítico às informações estatísticas na web (em forma de cubos para análise). Esta abordagem segue os moldes de SaaS (Software as a Service) ou seja o usuário final pode ter o serviço sem ter que criar e manter uma infraestrutura local para essa análise. O uso de várias tecnologias avançadas guiadas por essa sistemática, visa contribuir com as melhores práticas para o governo eletrônico, bem como a efetiva disponibilização de um protótipo para análise dos dados da PNAD em forma de serviço na web.

Palavras-Chave

Data Warehousing; Governo Eletrônico; SaaS; PNAD

Abstract

Rabelo, Flávia Andrade; Melo, Rubens Nascimento (Advisor). **A Systematic Based on Data Warehousing to Support the e-Government.** Rio de Janeiro, 2010. 123p. MSc. Dissertation – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro

Aiming to adapt to the new way of relationship between government and society, governments are faced with the growing need to offer easy access to public information to a public that increases everyday. To contribute to this purpose, the main objective of this paper is to develop a system based on Data Warehousing to support the electronic government. Thus, the objective is to offer a service of access to the analytical information to society via Internet and low cost. As a proof of concept of the systematical proposal, it will be used real data from a statistical service (PNAD), available on the web. The systematic application of these data generates an analytical access service to information on the web (in cubes for analysis). This approach follows the SaaS pattern (Software as a Service), that is, the user may use this service without having to create and maintain local infrastructure for this analysis. The use of several advanced technologies guided by that systematic, aims to contribute to best practices for e-government as well as the effective provision of a "demo" version of a cube for PNAD data analysis as a service on the web.

Keywords

Data Warehousing; e-Government; SaaS; PNAD

Sumário

1	Introdução	14
1.1.	Objetivos	18
1.2.	Delimitações da Pesquisa	18
1.3.	Motivação	19
1.4.	Contribuições Esperadas	19
1.5.	Organização da Dissertação	20
2	Fundamentação Teórica	22
2.1.	Governo Eletrônico	22
2.2.	Transparência	24
2.3.	Estatística Pública	26
2.4.	OLAP	28
2.5.	Data Warehousing	32
2.5.1.	Modelagem Dimensional	35
3	Sistemática Proposta	40
3.1.	Cenário de Aplicação	41
3.2.	Formalização do Problema	42
3.2.1.	Etapa 1 – Modelos conceituais das estatísticas	43
3.2.2.	Etapa 2 - Especificação dos requisitos OLAP	45
3.2.3.	Etapa 3 – Modelo lógico do Data Warehouse	47
3.2.4.	Etapa 4 - Metadados	48
3.2.5.	Etapa 5 – A Arquitetura	49
3.2.6.	Etapa 6 – Projeto de ETL e Desenvolvimento	50
3.2.7.	Etapa 7 – Disponibilizar os dados na Web	52
4	Aplicação da Sistemática	53
4.1.	Etapa 1 – Modelo Conceitual da Estatística	53
4.2.	Etapa 2 – Especificação de Requisitos OLAP	56
4.3.	Etapa 3- Construção da Modelagem Dimensional	58
4.4.	Etapa 4 - Metadados	58

4.5. Etapa 5- Definição da Arquitetura	60
4.6. Etapa 6 - Projeto de ETL e desenvolvimento	61
4.7. Etapa 7 – Acesso via Web	68
5 A Usabilidade das Estatísticas Públicas	71
5.1. Comparação do processo atual com o proposto	71
5.2. Usabilidade da Informação estatística em formato cúbico	73
5.3. Análise Exploratória da PNAD após a sistemática	74
6 Trabalhos Correlatos	80
6.1. Abordagem do governo americano	80
6.2. Abordagem do governo britânico	82
6.3. Abordagem do governo brasileiro	83
6.4. Considerações finais	85
7 Conclusão e Trabalhos Futuros	87
7.1. Conclusões	87
7.2. Trabalhos futuros	89
7.2.1. Intervalo de Confiança em Pesquisas Amostrais	89
7.2.2. Tecnologia “Open Source”	89
7.2.3. Entrega de dados geográficos	90
8 Referências Bibliográficas	91
Anexo I	98

Lista de figuras

Figura 1 - Websites com interatividade para acesso as informações governamentais, outubro 2009 (UN-Survey, 2010)	16
Figura 2 - Degraus da Transparência (Cappelli & Leite, 2008)	24
Figura 3 – Arquitetura de DWing (Fonte: autor)	32
Figura 4 – Exemplo de um esquema multidimensional	35
Figura 5 – Exemplo esquema lógico dimensional estrelado	39
Figura 6 – Exemplo de esquema físico dimensional estrelado	39
Figura 7 – Resumo da sistemática (fonte: a autora)	43
Figura 8 – Esquema dimensional conceitual de uma estatística pública	45
Figura 9 – Agrupamento de variáveis	46
Figura 10 – Exemplo de esquema lógico dimensional	48
Figura 11 - Arquitetura da sistemática proposta	50
Figura 12 – Esquema dimensional resultante	56
Figura 13 - Definição das estruturas das tabelas dimensões	57
Figura 14 - Esquema lógico Dimensional da PNAD Básica	58
Figura 15 - Modelo Físico metadados semânticos	59
Figura 16 - Diagrama da arquitetura na aplicação da sistemática PNAD Básica	61
Figura 17 – Trecho do arquivo “Pes2008.txt”	62
Figura 18 – Modelo físico das fontes de dados importados para o ODS	64
Figura 19 - Pessoas que freqüentavam creche ou escola, segundo a rede de ensino - Brasil – 2008	68
Figura 20 - População residente, por sexo, segundo os grupos de idade - Brasil – 2008	68
Figura 21 - Aplicação Web “open source“ OpenI para acesso OLAP	69
Figura 22 - Aplicação “open source” Palo Web Cliente para acesso OLAP	70
Figura 23 - Diagrama de atividades em cenários sem a aplicação da sistemática proposta	72
Figura 24 - Diagrama de atividades em cenários, após a aplicação da sistemática proposta	73
Figura 25 - População total do Brasil segundo faixa etária, 2007 e 2008	76
Figura 26 – Faixa de Renda per capita familiar do Brasil, 2007 e 2008	77
Figura 27- – Faixa de renda segundo a cor, Brasil – 2007	79

Figura 28 - Ferramentas para acesso e manipulação de dados do site data.gov.	81
Figura 29 – Abordagem do governo britânico	83
Figura 30 - Sac SP- Sistema de estatísticas das reclamações feitas a prefeitura de SP	85
Figura 31 - Sistema que busca as linhas de ônibus na cidade de São Paulo	85

Lista de tabelas

Tabela 1 – Definição de Graus (Cappelli, 2009)	25
Tabela 2 - Formato do arquivo de metadados da PNAD	62
Tabela 3 - Formato do arquivo de metadados após padronização	63
Tabela 4 - Tabela pess2008 correspondente aos dados da PNAD	64
Tabela 5 - Dicionário de variáveis correspondente aos metadados da PNAD	64
Tabela 6 - Distribuição da população segundo faixa etária	76
Tabela 7 – Faixa de renda per capita familiar – população total	77
Tabela 8 – Faixa de renda segundo a cor, Brasil – 2007	78

Lista de siglas e abreviaturas

BDM	Banco de Dados Multidimensional
BI	Business Intelligence (inteligência de negócio)
DDL	Data Definition Language
DML	Data Manipulation Language
DW	<i>Data Warehouse</i> (Armazém de dados)
DWing	Data Warehousing (Processo de construção de DW)
ePing	Programa de Interoperabilidade do Governo
ETL	<i>Extract Transform Load</i> (Extração Transformação Carga)
FOIA	<i>Freedom of Information Act</i>
IBGE	Instituto Brasileiro de Geografia e Estatísticas
INEP Teixeira	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
MDX	MultiDimensional eXpression(Linguagem de consulta multidimensional)
OCDE	Organização de Cooperação para o Desenvolvimento Econômico
OLAP	On-line Analytical Processing (Processamento Analítico On-Line)
OLTP	On Line Transaction Process(Sistemas de transação)
OMS	Organização Mundial de Saúde
ONG	Organização Não Governamental
ONU	Organização das Nações Unidas
PNAD	Pesquisa Nacional por Amostra de Domicílios
SaaS	<i>Software as a Service</i> (Software como um Serviço)
SEADE	Sistema Estadual de Análise de Dados
SEN	Sistema Estatístico Nacional
SGBD	Sistema Gerenciador de Banco de Dados
SGBDM	Sistemas Gerenciadores de Banco de Dados Multidimensionais
SQL	<i>Structured Query Language</i> (Linguagem
XML	<i>eXtensible Markup Language</i> (Linguagem de marcação de texto)
XMLA	<i>XML for Analysis</i> (Linguagem de marcação de texto para Análises)
WOLAP	Web Olap