



Universidade Estadual de Campinas  
Faculdade de Engenharia Elétrica e de Computação  
Departamento de Comunicações



**Uma Proposta de Utilização das Tecnologias de *Business Intelligence*  
para Suporte a Tomada de Decisão no Contexto de Governo  
Eletrônico**

**Autor: Eduardo Zanoni Marques**

Orientador: Prof. Dr. Leonardo de Souza Mendes

Dissertação de Mestrado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para a obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: **Telecomunicações e Telemática.**

Banca Examinadora

Leonardo de Souza Mendes, Dr. (Presidente)  
Mario Lemes Proença Jr., Dr.  
Rodolfo Miranda de Barros, Dr.

DECOM/FEEC/UNICAMP  
DC/UEL  
DC/UEL

Campinas, SP  
31 de Março de 2011

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

M348p Marques, Eduardo Zanoni  
Uma proposta de utilização das tecnologias de business intelligence para suporte a tomada de decisão no contexto de governo eletrônico / Eduardo Zanoni Marques. --Campinas, SP: [s.n.], 2011.

Orientador: Leonardo de Souza Mendes.  
Dissertação de Mestrado - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Governo eletrônico. 2. Sistemas de suporte de decisão. 3. Business intelligence. I. Mendes, Leonardo de Souza. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Título em Inglês: A proposal for business intelligence technologies utilization to decision making support in the electronic government context

Palavras-chave em Inglês: Electronic government, Decision support systems, Business intelligence

Área de concentração: Telecomunicações e Telemática

Titulação: Mestre em Engenharia Elétrica

Banca examinadora: Mario Lemes Proença Junior, Rodolfo Miranda de Barros

Data da defesa: 31/03/2011

Programa de Pós Graduação: Engenharia Elétrica

## COMISSÃO JULGADORA - TESE DE MESTRADO

**Candidato:** Eduardo Zanoni Marques

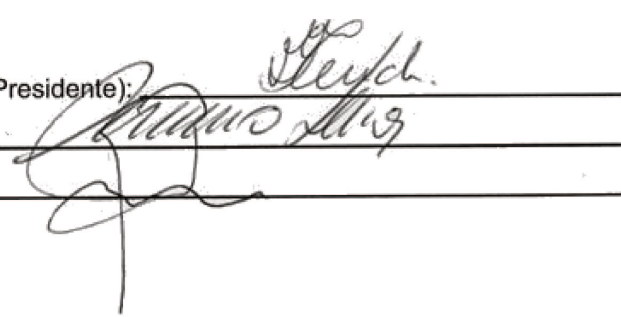
**Data da Defesa:** 31 de março de 2011

**Título da Tese:** "Uma Proposta de Utilização das Tecnologias de Business Intelligence para Suporte a Tomada de Decisão no Contexto de Governo Eletrônico"

Prof. Dr. Leonardo de Souza Mendes (Presidente):

Prof. Dr. Mario Lemes Proença Junior:

Prof. Dr. Rodolfo Miranda de Barros:



The image shows three horizontal lines representing signature lines. The first line has a handwritten signature that appears to be 'Leonardo S. Mendes'. The second line has a handwritten signature that appears to be 'Mario Lemes Proença Junior'. The third line has a handwritten signature that appears to be 'Rodolfo Miranda de Barros'. The signatures are written in black ink and are somewhat stylized.

## Resumo

A análise e interpretação dos dados gerados pelos sistemas de informação é uma realidade para as organizações públicas e privadas. Para atender a esta necessidade, surgiram os Sistemas de Suporte à Decisão. Dentre estes, os sistemas de *Business Intelligence* foram e estão sendo adotados pelo setor privado. No entanto, poucos trabalhos discutem a adoção destes sistemas para o setor público. Este trabalho discute as formas como os sistemas de *Business Intelligence* podem ser aplicados para auxiliar a administração pública, considerando as particularidades existentes neste setor. Ainda, este trabalho apresenta uma proposta pra construção de um ambiente de *Business Intelligence* utilizando ferramentas e tecnologias de código aberto. Este ambiente foi aplicado em um cenário real, onde é feita análise dos dados sociais da cidade de Campinas, SP.

**Palavras-chave:** Governo Eletrônico, Sistema de Suporte à Decisão, *Business Intelligence*.

## Abstract

The analysis and interpretation of data generated by information systems is a reality for both public and private organizations. As a solution to this problem, the Decision Support Systems were created. Among those, Business Intelligence systems were and are being adopted in the private organizations. Even though, just a few papers discuss the adoption of these systems in public organizations. This dissertation discusses how Business Intelligence systems can be applied to support public administration, considering the public organizations particularities. Still, it presents a proposal to create a Business Intelligence environment using open source tools and technologies. This environment was applied in a real scenario, where the social data from the city of Campinas, SP, was analyzed.

**Keywords:** Electronic Government, Decision Support System, Business Intelligence.

## Agradecimentos

Inicialmente, agradeço ao meu Deus, pela força e sabedoria que tem me dado durante todos os anos de minha vida e que me permitiram chegar até aqui e vencer mais um desafio da vida.

Agradeço à minha família por todo o apoio dado durante os anos, acreditando e tendo paciência comigo, sendo sempre meu porto seguro. Em especial, agradeço minha irmã Marcia pelos inúmeros puxões de orelha, revisões e consultorias prestadas no decorrer deste trabalho. Valeu maleta! Sem você nada disto teria acontecido.

Outra pessoa que devo demais é a minha menina Dani. Obrigado por estar ao meu lado na fase mais difícil deste trabalho, pelas correções incisivas que contribuíram tanto para a qualidade literária da dissertação, por me acalmar e me trazer alegria nos dias mais difíceis. Enfim, obrigado por estar sempre presente. Você é umas das principais culpadas pela conclusão deste trabalho.

Agradeço ao meu grande amigo e “co-orientador” Miani, que tanto me ajudou para organização das ideias e me deu confiança no desenvolvimento do trabalho. Faltou você pra comer uns pintxos lá em Bilbao.

Agradeço aos amigos Gean e Zarpelão por toda ajuda na fase final do trabalho, se tornando atores elementares nesta fase. Obrigado de verdade!

Agradeço ainda aos companheiros de laboratório do projeto Harpia e SIGM. A amizade e companheirismo de vocês foram essenciais para que pudesse vencer esta barreira e tantas outras que nos foram impostas. De todos os grandes amigos, agradeço em especial: ao menino Marlinho e Ékler (o 36 é o mais legal!!!); Derek, Burga e Andoré, amigos de longa data; Dylon, Cubas, Henrique, Sergião, Camila, Bruno Tx, Daniboy, Prof, Kadu, Tânia, Livia e Gabi, a galera do projeto Harpia; e Éverton (o homem da ideia), André, Kbelo, Líniquer, Alexandre, Richard, Tilli e Éder, a galera do SIGM.

Agradeço aos amigos Corvo, Michael e Vinícius que, mesmo de longe, me ajudam e me apoiam há tantos anos. Saibam que amizade de vocês é especial para mim.

Agradeço também ao companheiro de república Erick, pelas noites de Libertadores e por todas as outras de conversas até altas horas, sempre com a companhia do Derek e do Laranja.

Por fim, agradeço ao meu orientador Leonardo pela oportunidade e por toda a ajuda prestada durante os anos de desenvolvimento deste trabalho. Obrigado pelas palavras sábias que tentei transferir para este trabalho e que certamente levarei comigo para toda a vida.

# Sumário

Lista de Figuras .....	x
Lista de Tabelas .....	xii
Glossário.....	xiii
1 Introdução.....	1
1.1 Sistemas de Suporte à Decisão .....	2
1.2 Objetivos.....	4
1.3 Organização da Dissertação.....	4
2 Sistemas de Governo Eletrônico .....	6
2.1 Introdução.....	6
2.2 Arquiteturas de e-Gov .....	8
2.2.1 Camada de Acesso.....	9
2.2.2 Camada de e-Gov .....	10
2.2.3 Camada de e-Business .....	11
2.2.4 Camada de Infraestrutura.....	11
2.3 Classificação dos Sistemas de e-Gov .....	12
2.3.1 G2C.....	13

2.3.2	G2B.....	14
2.3.3	G2G .....	14
2.3.4	IEE.....	15
2.3.5	Infraestrutura Global.....	15
3	<i>Business Intelligence</i> .....	16
3.1	Processo de Tomada de Decisão .....	16
3.2	Tecnologias Envolvidas.....	18
3.2.1	Obtenção e Padronização dos Dados.....	18
3.2.2	Armazenamento dos Dados .....	21
3.2.3	Aplicativos dos Usuários Finais .....	25
3.2.3.1	Aplicativos de Consulta e Relatório.....	26
3.2.3.2	Aplicativos OLAP.....	26
3.2.3.3	Aplicativos de Mineração de Dados.....	30
3.3	Criação de Ambientes de BI.....	32
4	Ambiente de <i>Business Intelligence</i> para e-Gov.....	34
4.1	Criação de Ambientes de BI para e-Gov .....	34
4.2	Aplicabilidade dos Ambientes de BI para e-Gov .....	35
4.3	Ambiente Proposto .....	36
4.3.1	Camada de ETL.....	37
4.3.2	Camada de Armazenamento e Disponibilização de Visões dos Dados.....	39



4.3.3	Camada de Aplicações para os Usuários Finais .....	42
4.3.4	Aspectos de Segurança do Ambiente .....	44
4.3.4.1	Segurança na comunicação de dados .....	45
4.3.4.2	Segurança no acesso aos dados .....	46
5	Estudo de Caso .....	48
5.1	Contexto de Aplicação da Proposta.....	48
5.2	Metodologia Adotada .....	49
5.3	Primeira Iteração.....	51
5.4	Segunda Iteração.....	54
5.5	Análise da Aplicação do Ambiente de BI .....	57
6	Conclusão .....	60
7	Referências Bibliográficas .....	63

## Lista de Figuras

Figura 2.1 - Arquitetura das aplicações de e-Gov.. .....	9
Figura 2.2 - Classes de aplicações de e-Gov.. .....	13
Figura 3.1 - Exemplo de cubo de dados representado no modelo estrela. ....	23
Figura 3.2 - Exemplo de cubo de dados representado no modelo floco de neve. ....	25
Figura 3.3 - Tabelas de resultado da análise das vendas. ....	28
Figura 3.4 - Exemplo da operação <i>rollup</i> . ....	29
Figura 3.5 - Exemplo da operação <i>drill-down</i> . ....	29
Figura 3.6 - Exemplo da operação <i>slice and dice</i> . ....	30
Figura 3.7 - Exemplo da operação <i>pivot</i> . ....	30
Figura 4.1 – a) Arquitetura proposta por Moss e Atre (2003) b) Ambiente de BI proposto.	37
Figura 4.2 - Exemplo de utilização da ferramenta Talend Open Studio. ....	38
Figura 4.3 - Ferramenta Mondrian Schema Workbench. ....	41
Figura 4.4 - Resultado de uma consulta no OpenI na forma de gráfico de barras. ....	42
Figura 4.5 - Resultado de uma consulta no OpenI na forma de tabela multidimensional. ....	43
Figura 4.6 - Soluções adotadas para a segurança da comunicação de dados. ....	45
Figura 5.1 - Diagrama dos cubos de dados candidatos da primeira iteração. ....	51
Figura 5.2 - Diagrama conceitual do cubo de dados <i>Inclusão Programa Social</i> . ....	52

Figura 5.3 - Diagrama físico do cubo de dados <i>Inclusão Programa Social</i> .....	52
Figura 5.4 - Consulta executada sob dados fictícios no ambiente do estudo de caso. ....	54
Figura 5.5 - Diagrama dos cubos de dados candidatos da segunda iteração. ....	55
Figura 5.6 - Modelo conceitual dos cubos de dados selecionados na segunda iteração.....	56
Figura 5.7 - Diagrama conceitual do cubo de dados <i>Atendimento</i> . ....	56
Figura 5.8 - Diagrama físico do cubo de dados <i>Atendimento</i> .....	57

## Lista de Tabelas

Tabela 3.1 - Comparação entre os sistemas OLTP e OLAP (Vercellis, 2009). .....	27
Tabela 3.2 – Resultado da análise multidimensional das vendas. ....	28
Tabela 4.1 - Comparativo entre as soluções para camada de ETL.....	39
Tabela 4.2 - Comparativo entre as soluções para camada de Visão Lógica dos Dados. ....	41
Tabela 4.3 - Comparativo entre as soluções de clientes OLAP. ....	43

## Glossário

BI	<i>Business Intelligence</i>
CIO	<i>Chief Information Officer</i>
CSV	<i>Comma Separated Value</i>
DDL	<i>Data Definition Language</i>
DSS	<i>Sistemas de Suporte à Decisão – Decision Support Systems</i>
DW	<i>Data Warehouse</i>
e-Gov	Governo Eletrônico – <i>Electronic Government</i>
ER	Entidade Relacionamento
ERP	Sistema Integrado de Gestão Empresarial – <i>Enterprise Resource Planning</i>
ETL	Extração, Transformação e Carga – <i>Extract, Transform and Load</i>
G2B	Governo para Empresas – <i>Government to Business</i>
G2C	Governo para o Cidadão – <i>Government to Citizen</i>
G2E	Governo para Funcionários – <i>Government to Employees</i>
G2G	Governo para Governo – <i>Government to Government</i>

GIS	Sistema de Informação Geográfica – <i>Geophysical Information Systems</i>
GSB	Barramento Governamental de Serviços – <i>Government Service Bus</i>
HTML	<i>HyperText Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
HTTPS	<i>Hypertext Transfer Protocol Secure</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IEE	Eficiência e Efetividade Interna – <i>Internal Efficiency and Effectiveness</i>
JDBC	<i>Java Database Connectivity</i>
MAN	Redes Metropolitanas – <i>Metropolitan Area Networks</i>
MDX	<i>Multi Dimensional Expressions</i>
OLAP	Processamento Analítico de Tempo Real – <i>Online Analytical Processing</i>
OLTP	Processamento de Transações em Tempo-real – <i>Online Transactional Processing</i>
Open MAN	Redes Metropolitanas de Acesso Aberto – <i>Open Access Metropolitan Area Networks</i>
PDI	Pentaho Data integration
PDF	<i>Portable Document Format</i>

PNAD	Pesquisa Nacional por Amostra de Domicílios
SGBD	Sistema de Gerenciamento de Banco de Dados
SIGM	Sistema Integrado de Governança Municipal – <i>Integrated System for Municipal e-Gov</i>
SOAP	<i>Simple Object Access Protocol</i>
SP	São Paulo
SQL	<i>Structured Query Language</i>
SSL	<i>Secure Socket Layer</i>
TI	Tecnologia da Informação
TIC	Tecnologias de Informação e Comunicação
TLS	<i>Transport Layer Security</i>
TOS	Talend Open Studio
UML	Linguagem Unificada de Modelagem – <i>Unified Modeling Language</i>
XML	<i>Extensible Markup Language</i>
XMLA	<i>XML for Analysis</i>

## Trabalhos Publicados Pelo Autor

1. MARQUES, Eduardo Zanoni, MIANI, Rodrigo Sanches, GAGO JÚNIOR, Everton Luiz de Almeida, MENDES, Leonardo de Souza, **Development of a *Business Intelligence* Environment for e-Gov using Open Source Technologies**, XII International Conference on Data Warehousing and Knowledge Discovery – DaWak 2010, Bilbao - Espanha 30/08-02/09/2010.



# 1 Introdução

Com o progresso tecnológico, a utilização de soluções de *software* vem aumentando para as mais diversas áreas. De forma geral, os sistemas de informação são focados no armazenamento e processamento de dados, que auxiliam o setor operacional das organizações a executar suas tarefas diárias como registrar vendas, calcular folhas de pagamento e registrar dados pessoais de funcionários e clientes.

Esta realidade, comum das organizações privadas, também está presente na esfera pública. A aplicação das tecnologias de informação e comunicação (TIC) neste contexto, definida como Governo Eletrônico (e-Gov), vem ocorrendo estimulada pelas demandas dos cidadãos por maior transparência e responsabilidade na gestão dos recursos públicos. Assim, os governos buscam com o e-Gov meios de inovar e aprimorar o relacionamento entre os órgãos públicos e o cidadão, auxiliar as organizações públicas na gestão de recursos e possibilitar o monitoramento do impacto da implantação de políticas públicas na sociedade, entre outros (Nações Unidas, 2008).

Dentro desta realidade, existe uma demanda eminente por soluções de *software* que auxiliem os setores gerenciais destas organizações, públicas e privadas, na análise e interpretação de seus dados operacionais, auxiliando na tomada de decisão. Assim, estas soluções permitem às organizações o acompanhamento mais claro dos seus setores, possibilitando identificar deficiências e/ou oportunidades de negócio (Xu *et al*, 2007).

Buscando sanar esta necessidade várias ferramentas, tecnologias e soluções foram criadas, as quais estão agregadas pelo conceito de Sistemas de Suporte à Decisão (do inglês *Decision Support Systems* – DSS) (Power, 2002).

## 1.1 *Sistemas de Suporte à Decisão*

O desenvolvimento das ferramentas de DSS teve seu início no fim dos anos 60 por meio da combinação de estudos teóricos sobre o processo de tomadas de decisões e de estudos práticos para criação de sistemas de computação interativos (Keen e Morton, 1978). Segundo Power (2002), estes sistemas têm como principais características:

1. são desenvolvidos especificamente para auxílio no processo de tomada de decisão;
2. devem auxiliar, e não automatizar, a tomada de decisão;
3. devem ser rápidos para atender às mudanças requeridas pelos responsáveis pela tomada de decisão.

Inicialmente, estes sistemas trabalhavam com modelos matemáticos para simulação de cenários de negócio. Com o tempo novas técnicas foram agregadas, ampliando o escopo de aplicações e recursos destes sistemas. Atualmente, os DSS podem ser classificados em cinco tipos distintos (Power, 2002):

- **DSS orientados a modelos:** focam-se no acesso e manipulação de modelos financeiros, de otimização e simulação, tendo tipicamente como base um conjunto restrito de parâmetros e dados;
- **DSS orientados a dados:** tem como objetivo o acesso e manipulação de dados internos das organizações em determinado período de tempo, podendo também trabalhar com dados externos e/ou de tempo real;
- **DSS orientados a comunicações:** buscam facilitar colaborações e comunicações relevantes para a tomada de decisão, utilizando tecnologias de comunicação com ferramentas como *groupware* e vídeo conferência;

- **DSS orientados a documentos:** visam auxiliar o processo de análise e busca de documentos, por meio do emprego de técnicas como indexação e busca de arquivos e imagens;
- **DSS orientados a conhecimento:** munidos de conhecimento sobre um domínio específico, estes sistemas fazem sugestões e recomendam aos gestores ações para a solução de problemas.

Dentre estes sistemas, os DSS orientados a dados apresentam grande disseminação entre as organizações, com destaque para as tecnologias de *Business Intelligence* (BI). A aplicação destas tecnologias nas organizações privadas vem trazendo resultados positivos (Aucoin, 2010). Outra evidência do sucesso destas aplicações é observada no relatório do *Gartner Group* de 2009, onde pelo quarto ano consecutivo as tecnologias de BI são descritas como a principal prioridade dentre as Tecnologias de Informação (TI) dos CIO (*Chief Information Officer*) (Gartner Group, 2009).

Embora existam diversas similaridades entre a aplicação de TIC nos ambientes públicos e privados, é preciso considerar fatores específicos que podem ser encontrados nos ambientes públicos, como regulamentações legais, inércia organizacional e cultural, infraestrutura e sistemas ultrapassados, baixo investimento nos setores de TI e carência de mão de obra especializada e treinada (Chen, 2002) (Ebrahim e Irani, 2005). Esta realidade dificulta a criação de um ambiente de BI, que exige pessoal treinado e capacitado para utilização das ferramentas e governança das TIC e dos dados pertencentes ao ambiente, além dos possíveis investimentos em treinamentos e na aquisição de licenças de *software* e equipamentos.

Visando abordar estes problemas, este trabalho discute como as aplicações de BI se encaixam no contexto de e-Gov e possibilitam a análise e interpretação dos dados produzidos pelos sistemas de e-Gov. Ainda, são propostos meios de contornar a questão do baixo orçamento dos departamentos e auxiliar na capacitação dos funcionários. Para tanto, são utilizadas tecnologias de código aberto que possuam

veículos livres para o aprendizado, troca de experiências e resolução de problemas e dúvidas.

## **1.2 Objetivos**

Este trabalho teve por objetivo geral analisar a viabilidade da aplicação de sistemas de suporte de decisão, em especial sistemas de BI, no auxílio à análise e interpretação dos dados operacionais de aplicativos focados no contexto de e-Gov. Para tanto, possui como objetivos específicos:

1. definir as principais características dos aplicativos de e-Gov e analisar suas especificidades;
2. levantar as características dos sistemas de BI e descrever suas aplicabilidades;
3. discutir as formas como os sistemas de BI podem ser empregados nos sistemas de e-Gov;
4. apresentar uma proposta de aplicação de BI em sistemas de e-Gov, utilizando tecnologias de código aberto;
5. aplicar a proposta elaborada em um cenário real e apresentar a análise dos resultados obtidos.

## **1.3 Organização da Dissertação**

Este trabalho é organizado em seis capítulos, que buscam apresentar o encadeamento lógico do assunto abordado.

No capítulo dois são descritos os sistemas de e-Gov, demonstrando suas características, arquitetura e classificação.

O capítulo três traz a revisão da literatura sobre aplicações de BI, tratando das formas como estes podem auxiliar no processo de tomada de decisão e de seus principais componentes.

No capítulo quatro são apresentadas aplicações das técnicas de BI em diferentes contextos. É então elaborada a discussão sobre o modo como as técnicas de BI podem ser utilizadas no contexto de e-Gov. Em resposta à discussão, é proposto um ambiente de BI para e-Gov.

No capítulo cinco é demonstrada a aplicação deste ambiente proposto em um cenário real. Concluindo o capítulo, é feita a análise sobre os pontos positivos e negativos do ambiente observados no decorrer da sua implantação.

No capítulo seis são feitas as considerações finais e propostas de trabalhos futuros que possibilitem continuar a pesquisa aqui apresentada.

## 2 Sistemas de Governo Eletrônico

Neste capítulo é apresentada a definição de Governo Eletrônico, as razões que impulsionaram a sua criação, as particularidades da aplicação das TIC e a arquitetura sobre a qual os sistemas de e-Gov são desenvolvidos.

### 2.1 Introdução

Uma tendência de reforma nas instituições públicas surgiu em diferentes países ao redor do mundo, estimulada principalmente pelas novas e crescentes expectativas dos cidadãos sob seus governantes. O sucesso das lideranças públicas é cada dia mais mensurado com base nos benefícios garantidos aos constituintes. Sejam as organizações privadas, as comunidades ou os cidadãos, todos exigem eficiência e responsabilidade na gestão dos recursos e objetividade na entrega de melhores serviços e resultados (Nações Unidas, 2008).

Frente a este novo cenário, diversos países vêm buscando revitalizar suas administrações públicas, tornando-as mais proativas, transparentes e eficientes. Para isso, os governos procuram introduzir inovações nas suas estruturas, procedimentos e formas de mobilizar e utilizar seus capitais humanos e recursos financeiros e tecnológicos. Neste contexto, a utilização das TIC, quando aplicadas corretamente, possui papel chave para criação de um ambiente de crescimento econômico e social, onde a realização destas metas seja possível (Nações Unidas, 2008).

O termo Governo Eletrônico (*Electronic Government* – e-Gov) foi atribuído ao uso das TIC nas instituições públicas. Sprecher (2000) define e-Gov de duas formas, sendo que na mais restrita, e-Gov é definido como a produção e disponibilização de serviços governamentais através das TIC. Já na mais ampla, e-Gov é definido por toda forma de utilização das TIC para simplificar e melhorar as transações entre o governo e outros atores, como os cidadãos, o setor privado e outras organizações governamentais.

A utilização dos sistemas de e-Gov é amplamente baseada na aplicação das TIC no comércio, definido como Comércio Eletrônico (Barzilai-Nahon e Scholl, 2010). Isto ocorre devido aos desafios comuns destes sistemas, ambos tendo como objetivo facilitar a troca de bens, serviços e informações entre duas ou mais partes. Além disto, há uma preocupação comum com a confiança dos usuários sobre os serviços e informações fornecidos, decorrente da falta da interação face a face deste tipo de aplicação (Carter e Bélanger, 2004).

É importante destacar as diferenças existentes entre estas classes de sistemas. Gordon (2002) destaca que o principal papel do governo é regulamentar a sociedade enquanto no comércio o objetivo é a propaganda e vendas. Assim, um dos principais papéis do governo eletrônico é de auxiliar o trabalho e a comunicação dos diferentes poderes (executivo, legislativo e judiciário) e órgãos governamentais.

A relação com o cidadão é outro fator a se considerar. As soluções de *e-commerce* têm os cidadãos como clientes que buscam adquirir um bem ou serviço. Este cliente pode optar por diferentes opções de mercado que oferecem o que se deseja, baseado em critérios como custo/benefício, credibilidade e facilidades na compra/contratação. Os clientes são encarados como meios de obtenção de lucro, podendo haver ofertas e descontos distintos de acordo com o perfil do cliente. Esta realidade não é válida para a relação do governo com o cidadão, dado que não existe concorrência para a prestação de serviços governamentais, havendo uma espécie de monopólio. Também não é permitido que haja distinção no tratamento e na oferta de opções entre os cidadãos (Ciborra, 2005).

Além das diferenças elencadas, a criação de sistemas de e-Gov pode enfrentar diferentes barreiras específicas deste contexto, como (Chen, 2002) (Ebrahim e Irani, 2005):

- inércia cultural e organizacional: organizações públicas são conhecidas pela ineficiência e resistência a mudanças, podendo tornar a mudança cultural mais desafiadora que a mudança tecnológica;

- regulamentações legais e governamentais: os processos governamentais são repletos de leis e regulamentações para tornar claros seus direitos e deveres, que, embora bem intencionados, podem inibir a inovação destes processos;
- infraestrutura e sistemas desatualizados: como resultado da diminuição do capital para os departamentos dos governos, vê-se uma defasagem na infraestrutura e nos sistemas, que muitas vezes não são baseados em soluções de *Internet*;
- isolamento tecnológico entre os departamentos: observa-se que os departamentos adquirem seus próprios equipamentos e aplicações para as suas necessidades específicas, gerando muitas vezes dificuldades para o compartilhamento de informações e integração entre os departamentos;
- falta de pessoal capacitado: dado os baixos orçamentos destinados aos departamentos de TI, torna-se difícil reter os funcionários capacitados e treinar os funcionários remanescentes destes departamentos.

## **2.2 Arquiteturas de e-Gov**

Para que haja o desenvolvimento do e-Gov dentro da sociedade, é necessário que sejam adotadas diferentes tecnologias e estratégias para geração e disponibilização de serviços. Neste sentido, é importante que um projeto de implantação do e-Gov seja baseado em uma arquitetura de referência.

A definição de uma arquitetura de e-Gov tem por objetivo estabelecer padrões, aplicações, modelos de negócio e regulamentações que visem facilitar a integração para a troca de serviços e informações entre os diferentes atores envolvidos (empresas, governo, cidadão). Outra vantagem da definição da arquitetura é possibilitar o



entendimento do processo de implementação e requisitos das TIC adotadas em cada ponto da arquitetura (Ebrahim e Irani, 2005).

Com o intuito de definir uma arquitetura genérica de e-Gov, tem-se uma ampla gama de propostas disponíveis na literatura. Como exemplo destas propostas, pode-se citar os trabalhos de Ebrahim e Irani (2005), Beer, Kunis e Runger (2006), Hofman (2008), Finger (2009), Yan e Guo (2010).

Embora estes trabalhos difiram nas tecnologias ou no nmero de camadas existentes em suas propostas,  possvel observar que todas as arquiteturas se encaixam na proposta de Ebrahim e Inari (2005), que  formada por quatro camadas: Camada de Acesso, Camada de e-Gov, Camada de e-Business e Camada de Infraestrutura (Figura 2.1).

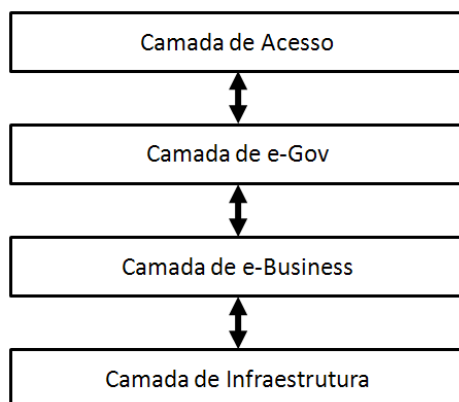


Figura 2.1 - Arquitetura das aplicaes de e-Gov. Adaptada de Ebrahim e Inari (2005).

### 2.2.1 Camada de Acesso

A Camada de Acesso tem como principal preocupao criar canais para que os servios, produtos e informaes disponibilizados pelo e-Gov possam ser utilizados, acessados e comunicados pelos cidados, governos e empresas. Estes meios consistem de canais *online*, como web sites que possam ser acessados via computadores e telefones celulares, e *offline*, como centrais de informao com funcionrios capacitados para auxlio na resoluo de problemas (Ebrahim e Inari, 2005).

Nesta camada, a principal questão a ser tratada é como estes canais podem permitir que diferentes usuários, com diferentes conhecimentos e interesses, tenham suas necessidades atendidas. É preciso que os canais possuam meios de atender usuários com diferentes competências técnicas, como capacidade de manipular o *mouse* e utilizar um *software*, e literacia da informação, que trata da capacidade do usuário em reconhecer que uma informação pode resolver seu problema ou atender a sua necessidade (Bélanger e Carter, 2009).

Outros pontos fundamentais que devem ser observados para o sucesso na realização desta camada são (Ebrahim e Inari, 2005):

- criação de mecanismos comuns para descoberta dos serviços, produtos e informações disponibilizadas;
- manutenção do sincronismo entre os canais de acesso;
- adoção de padrões de navegação e apresentação das informações comuns para os canais.

### **2.2.2 Camada de e-Gov**

Na Camada de e-Gov estão alocados todos os serviços oferecidos pelo governo. Para facilitar o acesso a todos os recursos disponibilizados através do e-Gov, é comum a criação de um portal *web* governamental. Este portal cria um ponto único de entrada para os usuários, facilitando a busca e utilização dos serviços (Ebrahim e Inari, 2005).

Outro benefício da criação do portal é a facilidade para o governo notificar os usuários sobre a inclusão de algum novo serviço ou a alteração de algum serviço já existente (Ebrahim e Inari, 2005).

### 2.2.3 Camada de e-Business

A Camada de e-Business é a primeira camada da arquitetura onde as TIC começam a ficar evidentes. Nela se encontram as soluções de *software* que possibilitam a informatização dos serviços governamentais. Como exemplo destas soluções tem-se os servidores de páginas *web*, sistemas de gerenciamento de banco de dados (SGBD) e sistemas integrados de gestão empresarial (do inglês *Enterprise Resource Planning* – ERP) (Ebrahim e Inari, 2005).

Uma das principais questões a ser tratada nesta camada é a integração dos dados e serviços dos diferentes órgãos e departamentos públicos. Como descrito na seção 2.1, estes órgãos e departamentos tradicionalmente utilizam bases de dados separadas e adotam sistemas e equipamentos para atender as suas necessidades particulares. Como resultado, torna-se um desafio complexo a criação de mecanismos para que os serviços prestados por eles se comuniquem, havendo a troca ou até o compartilhamento de informações (Yan e Guo, 2010).

### 2.2.4 Camada de Infraestrutura

Na Camada de Infraestrutura se encontram as soluções de *hardware* que permitem a disponibilização e entrega dos serviços governamentais aos usuários finais. Dentre estas soluções tem-se as máquinas onde estão instalados os aplicativos da camada de e-Business e placas de rede, roteadores e outros equipamentos que possibilitam a exposição dos serviços através da *Internet* ou de *Intranets* e *Extranets* (Ebrahim e Inari, 2005).

Com relação à exposição dos serviços aos usuários, é importante observar que o uso da *Internet* ainda é limitado a uma parcela da sociedade (Bélanger e Carter, 2009). Esta realidade é ilustrada pela Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2009, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), a qual

indica que apenas 41,7% dos entrevistados haviam utilizado a *Internet* nos últimos três meses (IBGE, 2009).

Uma das causas do pouco uso da *Internet* deve-se à ausência do acesso a *Internet* nos lares do cidadão. Segundo o PNAD de 2009, embora no Brasil 34,7% dos domicílios possua computador, apenas 27,4% deles está conectado a *Internet* (IBGE, 2009). Desta forma, é importante que os governos adotem medidas para expansão do acesso à *Internet* aos usuários do e-Gov.

Uma alternativa adotada pelos governos para resolução deste problema é a criação de Redes Metropolitanas de Acesso Aberto (do inglês *Open Access Metropolitan Area Networks* – Open MAN). As Open MAN ou Infovias são definidas por Mendes, Bottoli e Breda (2009) como uma rede multimídia convergente que, por meio da utilização de diferentes tecnologias de comunicação de dados, como a Ethernet com e sem fio, oferece acesso a todas as pessoas de um município. Por esta rede, é possível a integração dos órgãos governamentais e o provimento do acesso aos serviços governamentais aos cidadãos.

### **2.3 Classificação dos Sistemas de e-Gov**

De acordo com os dados apresentados até aqui, pode-se observar que os sistemas de e-Gov abrangem uma vasta área de aplicações, com diferentes objetivos e usuários finais.

De modo a organizar e agrupar as aplicações de e-Gov, Lee, Tan e Trimi (2005) discutem uma classificação para o e-Gov. Semelhante ao Comércio Eletrônico, esta classificação utiliza como fator de agrupamento das aplicações os usuários finais. Assim, são elencadas cinco classes de sistemas de e-Gov: Governo para o Cidadão (*Government to Citizen* – G2C), Governo para Empresas (*Government to Business* – G2B), Governo para Governo (*Government to Government* – G2G), Aplicações de Eficiência e Efetividade Interna (*Internal Efficiency and Effectiveness* – IEE) e

Aplicações de Infraestrutura Global (*Overarching Architecture*). Estas classes estão ilustradas na Figura 2.2.

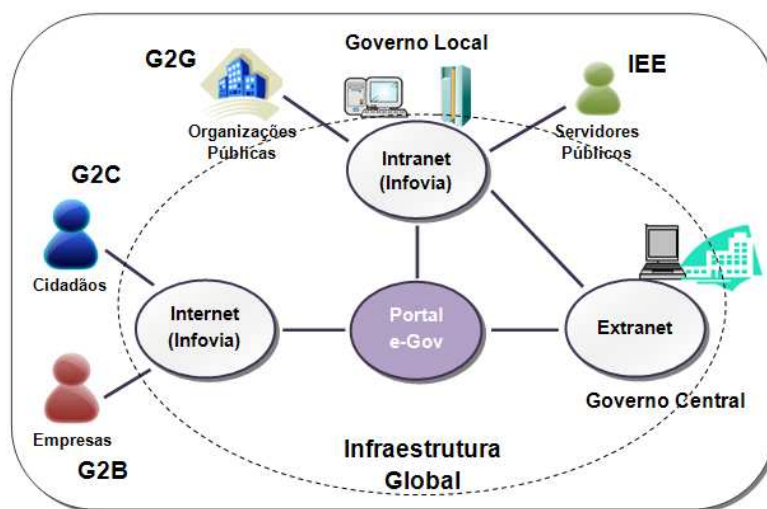


Figura 2.2 - Classes de aplicações de e-Gov. Adaptada de Yu (2009).

### 2.3.1 G2C

Na classe de G2C estão agrupadas as aplicações cujo usuário final são os cidadãos, as quais são as mais comuns nos ambientes de e-Gov. Através delas, são criados mecanismos para que os usuários possam, por exemplo, acessar as informações governamentais e dar as suas opiniões, além de solicitar serviços, como a poda de uma árvore da sua rua (Marchionini, Samet e Brandt, 2003).

Um exemplo de G2C é apresentado no trabalho de Velsen *et al* (2008), onde o autor discute a criação de um portal para cidadãos com necessidades especiais na Holanda. Uma proposta similar de criação de portal para o cidadão é apresentada no trabalho de Panhan, Ignatowicz e Mendes (2009), onde é descrita uma arquitetura para criação destes portais.

Outro trabalho relevante é o de Galpaya, Samarajiva e Soysa (2007). Nele, os autores analisam o baixo alcance da *Internet* para os cidadãos da Índia, Sri Lanka, Paquistão, Filipinas e Tailândia e propõem a utilização de centrais telefônicas para

provimento dos serviços governamentais aos cidadãos destes países, visto a maior disponibilidade desta tecnologia.

### **2.3.2 G2B**

A classe G2B contém aplicações cujo objetivo é aprimorar a comunicação entre governo e empresas. Através destas é possível que seja diminuída a carga burocrática imposta pelas agências governamentais nas relações de fornecimento de serviços e/ou mercadorias das empresas para os governos e de regulamentação governamental das atividades destas empresas (Lee, Tan e Trimi, 2005).

Um exemplo de sistema de G2B é descrito no trabalho de Dash e Sethi (2009), onde é apresentado como o governo da Índia criou mecanismos para controle de chamadas indesejadas de *telemarketing*. Foi criado um banco de dados central de telefones de usuários que não tinham interesse em receber estas chamadas. Para as empresas de *telemarketing* foi criado um serviço pelo qual estas submetiam suas listagens de telefone e lhes era retornado os números para os quais são permitidas as ligações.

### **2.3.3 G2G**

As aplicações que pertencem à classe G2G buscam a integração de diferentes órgãos e departamentos dos governos, promovendo meios de facilitar a troca e compartilhamento de informações e a integração dos serviços providos por estes (Lee, Tan e Trimi, 2005).

O trabalho de Dash e Sethi (2009) é também um exemplo de aplicação de G2G, pois de forma similar às empresas privadas de *telemarketing*, os serviços de *telemarketing* governamentais utilizam o serviço para verificar para quais telefones lhes é permitido efetuar ligações.

### **2.3.4 IEE**

Na camada de IEE estão as aplicações de e-Gov que buscam melhorar a eficiência de uma organização governamental. Isto pode ocorrer através da informatização e automação de processos internos e/ou da integração de diferentes sistemas pré-existentes na organização. Dentro desta classe está a classe de aplicações de Governo para Funcionários (*Government to Employees – G2E*) (Lee, Tan e Trimi, 2005).

O trabalho de Wagner e Antonucci (2004) é um exemplo das aplicações pertencentes a esta classe. Nele é discutido o processo de criação e implantação de um sistema de ERP para gestão da saúde pública no estado da Pensilvânia, Estados Unidos.

### **2.3.5 Infraestrutura Global**

As aplicações pertencentes à classe de Infraestrutura Global tratam de questões de interoperabilidade das aplicações pertencentes ao e-Gov. Esta questão pode ser abordada através de soluções de *hardware* e/ou *software* (Lee, Tan e Trimi, 2005).

Como exemplo deste tipo de aplicação, pode-se citar o trabalho de Yan e Guo (2010), onde é discutida a utilização de um Barramento Governamental de Serviços (*Government Service Bus – GSB*) para a integração e melhoria dos atributos de qualidade dos serviços *web* governamentais.

### **3 Business Intelligence**

O termo BI, criado por Howard Dresner, se refere à integração de tecnologias que vem se desenvolvendo ao longo de décadas para dar suporte ao processo de tomada de decisão (Martens, 2006). Neste capítulo será feita uma introdução ao processo de tomada de decisão, definindo o conceito de BI, as formas como este auxilia o processo de tomada de decisão e as tecnologias relacionadas ao mesmo.

#### **3.1 Processo de Tomada de Decisão**

Toda organização que disponibiliza os produtos e/ou serviços requeridos no local certo, no tempo certo e a um preço apropriado terá grande probabilidade de sucesso nos negócios. O problema, no entanto, é descobrir se o que está sendo oferecido é requerido naquele momento no local onde a oferta é feita e se o preço é o correto para aquele público. Para isso, é necessário que os responsáveis por tomar decisões sejam munidos de informações que as auxiliem a tomada das decisões corretas.

Neste sentido, podem-se observar dois problemas complementares que devem ser abordados para tornar o processo de tomada de decisão eficaz. É preciso criar formas de oferecer as informações aos tomadores de decisão em tempo hábil e avaliar se as decisões foram realmente corretas.

Inicialmente, é necessário definir o conceito de decisão correta, que pode ser definida como uma decisão que aproxima as organizações dos seus objetivos de negócio. Assim, para avaliar se uma decisão resulta nesta aproximação é preciso que as organizações definam metas a serem buscadas e formas de mensurar se, com o passar do tempo, essas organizações estão se movendo em direção favorável ou contra as metas estipuladas. Além disso, é essencial que sejam criados meios para que os dados gerados nesse processo sejam disponibilizados aos tomadores de decisão em tempo



hábil. Desta forma, esses dados serão utilizados tanto como fonte de informação para auxiliar a tomada de decisão como indicadores do ambiente, mostrando o resultado de uma decisão para a organização (Larson, 2009).

Para auxiliar o processo de coleta, organização, processamento e disponibilização destas informações, sistemas de *software* vêm sendo desenvolvidos, integrados e evoluídos, resultando nos atuais ambientes de BI.

Ambientes de BI tem como objetivo primário a melhoria do acesso e disponibilização dos dados de negócio aos tomadores de decisão das organizações. Desta forma, esses ambientes capacitam as organizações a tomarem decisões mais acertadas num espaço de tempo menor (Luckevich, Misner e Vitt, 2008).

Estes ambientes tipicamente são formados pela mistura de ferramentas, bases de dados e sistemas operacionais de diferentes fornecedores, criando uma infraestrutura que não só atende as necessidades iniciais das organizações, mas possui robustez para evoluir e adequar às necessidades futuras das organizações (Biere, 2003). Nestes ambientes é comum encontrar ferramentas de Processamento Analítico de Tempo Real (do inglês *Online Analytical Processing* - OLAP) (capítulo 3.2.3.2) e Mineração de Dados (do inglês *Data Mining*) (capítulo 3.2.3.3), além de outras de uso específico que fazem uso das capacidades destas ferramentas (Almeida *et al*, 1999).

Os ambientes de BI não produzem apenas fatos e figuras em um relatório. Estes sistemas trabalham buscando formas de promover a transformação consciente e metódica dos dados de todas as possíveis fontes, gerando informações orientadas ao negócio e ao resultado (Biere, 2003). Embora relatórios sobre processos ou atividades de uma organização sejam importantes, eles não são considerados BI até que estes dados sejam organizados de uma forma que as informações contidas sejam passadas de forma clara e direta aos tomadores de decisão em tempo para que essas informações realmente influenciem suas decisões. Desta forma, estes ambientes promovem uma mudança nas organizações, que deixam de ser reativas para se tornar proativas (Larson, 2009).

Outra preocupação destes ambientes é que as transformações dos dados sejam feitas da forma mais automática possível e que estes sejam disponibilizados aos tomadores de decisão sem que se criem dependências com soluções de um determinado fornecedor (Biere, 2003).

## **3.2 Tecnologias Envolvidas**

Como descrito, os ambientes de BI são criados a partir da integração de diferentes tecnologias. Estas tecnologias destinam-se a resolver problemas particulares do ambiente, as quais podem ser divididas em três grandes áreas: (1) obtenção e padronização dos dados de negócio, (2) armazenamento dos dados do BI e (3) aplicativos de *software* destinados aos usuários finais para acesso e manipulação dos dados do BI (Zeng *et al*, 2006).

### **3.2.1 Obtenção e Padronização dos Dados**

É comum que toda organização que deseja implantar um ambiente de BI certamente possui algum tipo de sistema de *software* que armazena integralmente ou parcialmente os dados gerados no seu dia a dia. Estes dados são armazenados em locais denominados repositórios de dados operacionais, que servem como principal fonte de dados para o ambiente de BI. Os dados destes repositórios podem estar armazenados de forma estruturada e não estruturada.

Repositórios de dados estruturados são aqueles cujos dados estão armazenados seguindo alguma estrutura lógica bem definida. Estes dados são geralmente mantidos por aplicativos de Processamento de transações em tempo-real (do inglês *Online Transactional Processing – OLTP*) utilizados pela organização. Os dados armazenados nestes repositórios são, em grande parte, de natureza numérica e se utilizam de estruturas como bancos relacionais, arquivos binários, arquivos no formato *Extensible Markup Language* (XML) e estruturas de dados hierárquicas (Kimball e Caserta, 2004).

Os sistemas de OLTP têm como principal preocupação o processamento e armazenamento de dados das transações de negócio que compõem o dia a dia das organizações. Estas transações são bem definidas, repetitivas, isoladas e de curta duração, requerendo dados de negócio atualizados que são tipicamente acessados através de chaves primárias. É necessário garantir que o processamento de uma nova transação não tornará os dados, previamente armazenados, inconsistentes e nem afetará no resultado de outras transações, sendo o controle ao acesso e alteração de dados concorrente uma das grandes preocupações destes sistemas (Chaudhuri e Dayal, 1997).

Nos repositórios de dados não estruturados estes estão armazenados de maneira casual, quase sem nenhum formato ou padrão, sendo originados por atividades informais dos usuários. A natureza dos dados neste formato é tipicamente de textos, normalmente armazenados em na forma de *emails*, arquivos de texto, planilhas e apresentações. Entretanto em cenários particulares, grande parte dos dados pode estar armazenada em imagens ou arquivos de fluxos de áudio e vídeo (Inmon, 2005).

Para a utilização destes dados no ambiente de BI é necessário que eles sejam submetidos a transformações que os ajustem às necessidades do ambiente de BI e armazenados em estruturas que facilitem o acesso destes pelas aplicações. Este processo é definido como Processo de Extração, Transformação e Carga (do inglês *Extract, Transform and Load* – ETL) (Kimball e Caserta, 2004).

O processo de ETL é composto por três passos sequenciais bem definidos: a extração, a transformação e a carga.

O passo da extração é responsável pelo acesso aos dados dos repositórios operacionais. Durante esta etapa, diversos desafios devem ser enfrentados, como estabelecer formas de acessar dados armazenados em estruturas particulares de uma aplicação ou em plataformas legadas como *mainframes*, criar meios de estruturar os dados contidos nos repositórios não estruturados e definir a codificação que será utilizada no acesso aos dados. Além disso, é comum usar filtros para definir quais dados serão extraídos, possibilitando diferenciar uma extração que tem como objetivo

fornecer os dados iniciais para um ambiente de BI de uma extração visando atualizar dados previamente existentes (Vercellis, 2009).

No passo de transformação, os dados extraídos pelo processo anterior passam pelos tratamentos necessários para que sejam adequados aos requisitos do ambiente de BI. Na execução deste passo o problema da qualidade dos dados contidos nos repositórios operacionais deve ser abordado. Dentre os principais problemas na qualidade temos: dados imprecisos com tamanhos e codificações diversificadas, incompletos ou inexistentes, inconsistências de padrões entre dados do mesmo tipo em repositórios distintos e desatualizados, redundantes e/ou irrelevantes (Vercellis, 2009).

Dentre os diferentes tipos de tratamentos aplicados aos dados para sua adequação tem-se (Kimball *et al*, 1998):

- correção de erros de digitação;
- ajuste de tamanho de campos;
- resolução de dados conflitantes;
- exclusão de valores desnecessários;
- combinação de dados complementares de diferentes repositórios;
- mapeamento de diferentes representações do mesmo dado para uma forma comum (como exemplo a informação sobre o sexo, que pode ser armazenada de diferentes formas, como F e M ou 0 e 1).

Por fim, no passo de carga, os dados que passaram pela fase de transformação são armazenados nos repositórios do BI. Durante esta etapa duas questões devem ser abordadas. A primeira é o tempo de atualização dos dados, visto que o tempo reservado para esta tarefa costuma ser bastante curto e em períodos de baixo ou nenhum uso do ambiente, como no período noturno. A segunda são os possíveis dados agregados existentes no ambiente, que devem ser recalculados, além de índices de dados que devem ser atualizados (Chaudhuri e Dayal, 1997).

Outras formas e funções do processo de ETL vêm sendo discutidas na literatura, como métodos para trabalhar o processo de ETL em tempo real e aprimorar a qualidade dos dados dos repositórios operacionais utilizando os dados corrigidos durante a etapa de transformação (Kalra e Steiner, 2005) e (Dayal *et al*, 2009).

### 3.2.2 Armazenamento dos Dados

Uma característica comum de todo ambiente de BI é que este possui um repositório próprio, contendo uma cópia unificada e padronizada pelo processo de ETL dos dados dos diferentes repositórios operacionais. As principais causas que justificam esta característica são (Inmon, 2005):

- as tecnologias que dão suporte às necessidades dos sistemas operacionais são diferentes das que dão suporte aos sistemas gerenciais e analíticos;
- a organização física dos dados utilizados pelos sistemas operacionais é diferente da utilizada pelos sistemas gerenciais e analíticos;
- as características de processamento dos sistemas operacionais é diferente das características dos sistemas gerenciais e analíticos;
- os usuários finais destas aplicações têm necessidades diferentes.

A principal solução de armazenamento de dados dos ambientes de BI é o *Data Warehouse* (DW). Este modelo de armazenamento foi inicialmente definido por Inmon (1992) como uma coleção de dados orientados a assunto, integrados, não voláteis e variantes ao tempo destinados a dar suporte às decisões gerenciais. A utilização do DW nos ambientes de BI é tão comum que Almeida *et al* (1999) e Biere (2003) consideram esses ambientes como a evolução do DW.

Desta forma, os dados no DW são armazenados e disponibilizados visando atender as necessidades de algum setor da organização (orientados a assunto). Nele, os

dados dos diferentes sistemas de informação da organização são centralizados, criando uma versão única da verdade de uma organização. Isso faz com que o problema de relatórios conflitantes desenvolvidos por departamentos diferentes de uma organização seja eliminado (integrados) (Biere, 2003). Estes dados são armazenados periodicamente (variantes ao tempo), sendo alterados apenas no caso de alguma correção (não voláteis). (Kimball e Ross, 2002).

Para a construção de um DW, encontramos na literatura duas propostas distintas sobre a sua estrutura, uma defendida por Inmon e outra defendida por Kimball.

Na visão de Inmon (2005), o DW deve possuir uma base de dados relacional centralizada, onde todos os dados devem ser armazenados. A partir dessa base são gerados *data marts* especializados para os departamentos dentro da organização. Desta forma, estes *data marts* são estruturas derivadas do DW e que podem ser descartadas sem causar nenhuma perda de dados ao DW.

Já na visão de Kimball (1996), o DW nada mais é que a integração de diversos *data marts*, que ele denomina *DW Bus Architecture*. Neste modelo, a integração dos *data marts* é feita através de dimensões similares (*conformed dimensions*), que contêm dados comuns entre diferentes cubos de dados, e fatos similares (*conformed facts*), que possuem significado e forma comuns entre diferentes *data marts*.

Independente da opção escolhida para implementação, uma das principais estruturas de um DW é o *data mart*. O *data mart* é definido como um conjunto de cubos de dados de um assunto específico voltado a atender as necessidades de sistemas gerenciais e analíticos. Os cubos de dados, exemplificado na Figura 3.1, são a representação mais simples com potencial analítico de um *data mart*, sendo compostos por duas entidades básicas, denominadas fato e dimensão (Kimball e Ross, 2002).

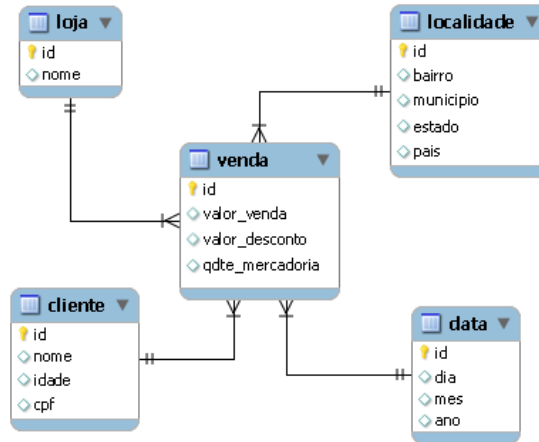


Figura 3.1 - Exemplo de cubo de dados representado no modelo estrela.

O fato é a entidade que representa os itens de interesse de uma organização. Neles estão inclusos os dados particulares do item, chamados métricas. Como regra geral, as métricas são valores numéricos, mas existem casos em que esta pode ser dada por valores textuais (Kimball e Ross, 2002). Um exemplo de fato são as vendas de uma loja (Figura 3.1), que tem como métricas o valor da venda, o valor do desconto dado e a quantidade de mercadorias vendidas.

As métricas podem ser classificadas como básicas e calculadas. As básicas são aquelas que têm seus valores consolidados durante o processo de ETL utilizando os dados dos repositórios operacionais. No caso das calculadas, seu valor é resultante da composição de outras métricas (Luckevich, Misner e Vitt, 2008).

A dimensão, por sua vez, é a entidade que contextualiza um fato. Nela estão armazenadas as descrições textuais de negócio, que permitem a análise dos dados de um fato sobre diferentes perspectivas (Kimball e Ross, 2002). Utilizando o exemplo do fato Venda, poderíamos ter como dimensão a loja onde foi feita a venda, a localidade da loja, o cliente que efetuou a compra e a data da compra (Figura 3.1).

Os dados da dimensão são organizados de forma hierárquica. Assim, as métricas podem ser analisadas em diferentes níveis de detalhamento, principal característica da análise multidimensional dos dados, descrita posteriormente no capítulo 3.2.3.2 (Kimball e Ross, 2002). Como exemplo de hierarquia de dados,

considere a dimensão *localidade*, da Figura 3.1. Nesta dimensão, os dados estão armazenados seguindo a hierarquia bairro → município → estado → país.

Uma questão a ser abordada no momento de definir um fato é a granularidade dos dados deste. A granularidade diz respeito ao nível mais baixo da hierarquia de uma dimensão o fato está relacionado. Fatos de granularidade mais baixa possuem maior potencial para análises, entretanto resultam em mais dados quando comparados a fatos de granularidade superior (Kimball *et al*, 1998).

Para a criação dos cubos de dados é utilizada a técnica da modelagem de dados multidimensional. Nesta modelagem, as principais preocupações são manter a estrutura dos dados mais próxima da forma compreendida pelos tomadores de decisão, facilitando a exploração dos dados e a geração de relatórios, e garantindo a velocidade na recuperação dos dados. Os dois principais modelos utilizados na modelagem multidimensional são o estrela e floco de neve, ambos criados por Ralph Kimball. Através destes é possível a criação de modelos de dados multidimensionais em bases de dados relacionais (Inmon, 2005).

No modelo estrela, a entidade fato é colocada no centro, sendo circundada pelas entidades dimensão, que estão relacionadas a ela. Assim, a disposição dos elementos do modelo lembra a figura de uma estrela (Kimball e Ross, 2002). Na Figura 3.1 tem-se um exemplo de cubo de dados no modelo estrela.

No modelo floco de neve, a disposição dos elementos é similar a do modelo estrela. No entanto, existe a separação dos dados de uma dimensão em duas ou mais entidades (Kimball e Ross, 2002). Um exemplo deste modelo pode ser visto na Figura 3.2, onde a dimensão *localidade* foi dividida nas entidades *loc\_municipio* e *loc\_estado\_pais*.



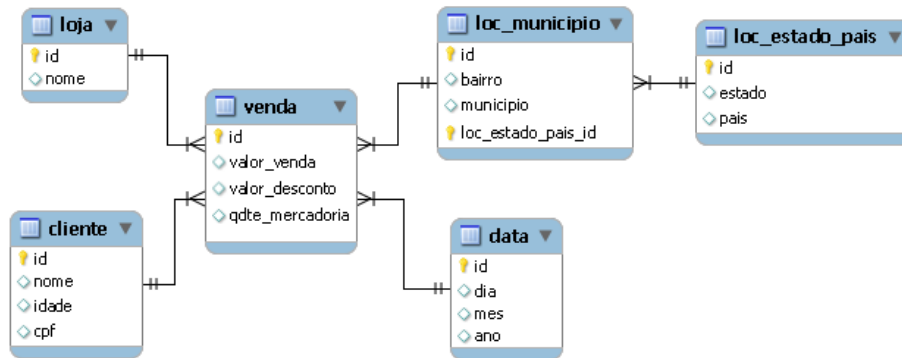


Figura 3.2 - Exemplo de cubo de dados representado no modelo floco de neve.

Os dois modelos apresentam vantagens e desvantagens. A normalização das dimensões promovidas pelo modelo floco de neve é de grande valia para a redução do espaço de armazenamento utilizado pelo DW, tendo em vista a diminuição de dados repetidos que devem ser armazenados. No entanto, esta mesma normalização faz com que haja o aumento do número de entidades do modelo, o que é prejudicial para a apresentação e compreensão do mesmo. Além disto, a separação dos dados faz com que o tempo de resposta do DW seja mais alto, visto o aumento no número de junções de entidades (Kimball e Ross, 2002).

### 3.2.3 Aplicativos dos Usuários Finais

Como dito anteriormente, o principal objetivo dos ambientes de BI é prover as informações aos tomadores de decisão de forma rápida e precisa. Com isso, pode-se observar que um dos pontos fundamentais do ambiente de BI são os aplicativos dos usuários finais, que são a porta de acesso dos usuários para as informações.

Apesar dos diferentes tipos de aplicativos que podem ser oferecidos aos usuários, define-se que as principais classes de aplicativos são os de Consulta e Relatório, OLAP e Mineração de Dados. Cada tipo de aplicativo possui uma função específica para diferentes tipos de usuário do ambiente, como descrito a seguir (Biere, 2003).

### **3.2.3.1 Aplicativos de Consulta e Relatório**

Esta classe de aplicativos tem por missão facilitar que o usuário navegue em dados armazenados nos mais diversos formatos e possa transformar o resultado desta navegação em relatórios que possam ser armazenados e facilmente acessados.

Segundo Biere (2003), algumas das características essenciais destes aplicativos são:

- acesso aos principais SGBD e formatos de arquivos mais comuns;
- interface de fácil utilização para o usuário;
- geração de consultas no padrão SQL;
- robustez para o controle de extensos resultados de consultas;
- oferecer múltiplos formatos para os relatórios, como PDF, HTML e Excel.

Os aplicativos de Análise e Relatório possuem maior alcance dentre os usuários finais do ambiente de BI, correspondendo a cerca de 80% destes usuários. Isso ocorre devido a esses aplicativos fornecerem dados básicos, utilizados para as decisões cotidianas de uma organização (Biere, 2003).

### **3.2.3.2 Aplicativos OLAP**

Buscando prover maiores possibilidades de análise aos dados, os aplicativos de OLAP foram criados. Estes disponibilizam um conjunto de mecanismos para a visualização e análise de dados de forma simples e rápida, utilizando tipicamente cubos de dados como fonte de dados. Este tipo de análise é denominado análise multidimensional de dados (Larson, 2009).

O termo OLAP e o conjunto inicial de funcionalidades destes aplicativos foram inicialmente definidos por E. F. Codd em 1993. A semelhança sintática entre os termos OLAP e OLTP é intencional, visando salientar as diferenças entre eles (Luckevich,

Misner e Vitt, 2008). A Tabela 3.1 apresenta um comparativo entre estes dois tipos de aplicativos.

Tabela 3.1 - Comparação entre os sistemas OLTP e OLAP (Vercellis, 2009).

Característica	OLTP	OLAP
Volatilidade	Dados dinâmicos	Dados estáticos
Natureza temporal dos dados	Apenas dados correntes	Dados correntes e históricos
Atualização	Contínua e irregular	Periódica e regular
Uso do aplicativo	Repetitivo	Imprevisível
Flexibilidade	Baixa	Alta
Desempenho	Alta, com diversas transações por segundo	Pode ser baixa para consultas complexas
Usuários	Funcionários comuns	Tomadores de Decisão
Funcionalidades	Operacionais	Analíticas
Propósito de uso	Transações	Consultas complexas e suporte à tomada de decisão
Prioridade	Alta performance	Alta flexibilidade
Tamanho	Megabytes a Gigabytes	Gigabytes a Terabytes

Para melhor entender o funcionamento e as vantagens da análise multidimensional dos dados, segue o seguinte exemplo descrito por Luckevich, Misner e Vitt (2008): um distribuidor de frutas deseja analisar os dados de distribuição dos dois últimos semestres. Este distribuidor trabalha com maçãs, uvas, bananas e morangos, que são revendidos para os estados de São Paulo, Goiás, Bahia e Paraná.

Ele inicia sua análise verificando a valor total vendido em cada um dos semestres (

Figura 3.3 – Tabela A). Percebendo que não houve diferenças entre os semestres, ele então decide analisar as vendas por produtos (

Figura 3.3 – Tabela B). Novamente ele não encontra diferenças, o que o faz analisar suas vendas com base nos estados (

Figura 3.3 – Tabela C). Em todas as análises, o resultado final sempre é de R\$32.000,00, o que sugere que em todas as análises estão sendo utilizados os mesmos dados de venda.

Período	Valor
1º Semestre	R\$ 16.000,00
2º Semestre	R\$ 16.000,00
<b>Total</b>	<b>R\$ 32.000,00</b>

Tabela A - Vendas Por Semestre

Estado	Valor
São Paulo	R\$ 8.000,00
Goiás	R\$ 8.000,00
Bahia	R\$ 8.000,00
Paraná	R\$ 8.000,00
<b>Total</b>	<b>R\$ 32.000,00</b>

Tabela B - Vendas por Estado

Estado	Valor
Maça	R\$ 8.000,00
Uva	R\$ 8.000,00
Banana	R\$ 8.000,00
Morango	R\$ 8.000,00
<b>Total</b>	<b>R\$ 32.000,00</b>

Tabela C - Vendas por Mercadoria

Figura 3.3 - Tabelas de resultado da análise das vendas.

Neste caso, todas as análises foram consideradas apenas uma dimensão, descrita no capítulo 3.2.2, para caracterizar as vendas. Na análise multidimensional, diferentes dimensões são utilizadas ao mesmo tempo para caracterizar um evento (no caso, a venda de frutas), possibilitando maior potencial para descoberta de padrões nos dados. Empregando esta análise nos dados do exemplo (Tabela 3.2), podem-se observar diversas características até então não observáveis, como a queda na venda de bananas e morangos nos estados de São Paulo e Goiás durante o segundo semestre, enquanto houve o aumento das vendas de maça e uva.

Tabela 3.2 – Resultado da análise multidimensional das vendas.

		São Paulo	Goiás	Bahia	Paraná	Total
<b>1º SEMESTRE</b>	<b>Maça</b>	-	-	R\$2.500,00	R\$1.500,00	R\$4.000,00
	<b>Uva</b>	-	-	R\$2.000,00	R\$2.000,00	R\$4.000,00
	<b>Banana</b>	R\$1.000,00	R\$3.000,00	-	-	R\$4.000,00
	<b>Morango</b>	R\$2.000,00	R\$2.000,00	-	-	R\$4.000,00
	<b>Total</b>	<b>R\$3.000,00</b>	<b>R\$5.000,00</b>	<b>R\$4.500,00</b>	<b>R\$3.500,00</b>	<b>R\$16.000,00</b>
<b>2º SEMESTRE</b>	<b>Maça</b>	R\$4.000,00	-	-	-	R\$4.000,00
	<b>Uva</b>	R\$1.000,00	R\$3.000,00	-	-	R\$4.000,00

<b>M E S T R E</b>	<b>Banana</b>	-	-	R\$1.500,00	R\$2.500,00	R\$4.000,00
	<b>Morango</b>	-	-	R\$2.000,00	R\$2.000,00	R\$4.000,00
	<b>Total</b>	<b>R\$5.000,00</b>	<b>R\$3.000,00</b>	<b>R\$3.500,00</b>	<b>R\$4.500,00</b>	<b>R\$16.000,00</b>
	<b>Total Anual</b>	<b>R\$8.000,00</b>	<b>R\$8.000,00</b>	<b>R\$8.000,00</b>	<b>R\$8.000,00</b>	<b>R\$32.000,00</b>

Aplicando as entidades da modelagem multidimensional discutidas no capítulo 3.2.2 neste exemplo, tem-se como fato a Venda, como métrica o Valor e como dimensões a Mercadoria com o atributo Nome, o Estado com o mesmo atributo e o Período com os atributos organizados na hierarquia semestre → ano.

Segundo Chaudhuri e Dayal (1997), as funcionalidades básicas de um aplicativo OLAP são: *rollup*, *drill-down*, *slice and dice* e *pivot*.

A operação de *rollup* permite que os dados sejam agregados a um nível superior da hierarquia onde a operação é aplicada

(Chaudhuri e Dayal, 1997). Um exemplo é apresentado na

	São Paulo	Goiás
Anual	R\$ 8.000,00	R\$ 8.000,00

Figura 3.4, onde os dados do fato Venda são agregados levando-se em conta a dimensão Período.

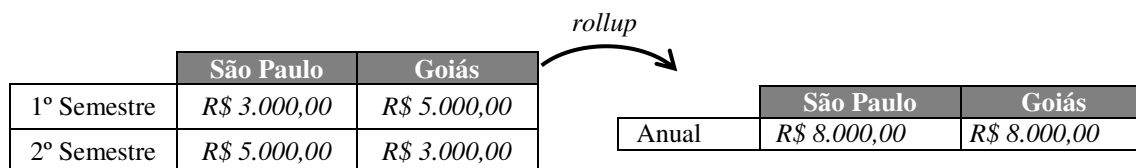


Figura 3.4 - Exemplo da operação *rollup*.

A operação *drill-down* tem função inversa à *rollup*, fazendo com que os dados apresentados sejam decompostos entre os filhos do nível inferior da hierarquia da dimensão onde a operação é aplicada (Chaudhuri e Dayal, 1997). Um exemplo é apresentado na Figura 3.5, na qual os dados agregados do fato Venda na dimensão Período são decompostos.

	São Paulo	Goiás
Anual	R\$ 8.000,00	R\$ 8.000,00

	São Paulo	Goiás
1º Semestre	R\$ 3.000,00	R\$ 5.000,00
2º Semestre	R\$ 5.000,00	R\$ 3.000,00

Figura 3.5 - Exemplo da operação *drill-down*.

A operação *slice and dice* permite que o usuário possa aplicar filtros sobre os dados consultados, gerando projeções destes (Chaudhuri e Dayal, 1997). Um exemplo dessa operação está ilustrado na Figura 3.6, onde os dados pré-consultados passam por uma operação de *slice and dice* que filtra o estado Goiás.

	São Paulo	Goiás
1º Semestre	R\$ 3.000,00	R\$ 5.000,00
2º Semestre	R\$ 5.000,00	R\$ 3.000,00

	São Paulo
1º Semestre	R\$ 3.000,00
2º Semestre	R\$ 5.000,00

Figura 3.6 - Exemplo da operação *slice and dice*.

Por fim, a operação *pivot* possibilita a reorientação da visualização multidimensional dos dados consultados (Chaudhuri e Dayal, 1997). Como exemplo dessa operação tem-se a

Figura 3.7, na qual a orientação horizontal e vertical das colunas é invertida.

	São Paulo	Goiás
1º Semestre	R\$ 3.000,00	R\$ 5.000,00
2º Semestre	R\$ 5.000,00	R\$ 3.000,00

	1º Semestre	2º Semestre
São Paulo	R\$ 3.000,00	R\$ 5.000,00
Goiás	R\$ 5.000,00	R\$ 3.000,00

Figura 3.7 - Exemplo da operação *pivot*.

### 3.2.3.3 Aplicativos de Mineração de Dados

Os aplicativos até aqui descritos têm como principal finalidade facilitar o acesso dos usuários aos dados armazenados no ambiente BI. No caso dos aplicativos

de mineração de dados, o objetivo é gerar informações escondidas em um conjunto de dados pré-existente, utilizando técnicas de mineração de dados. Estas informações devem possuir um significado que leve a alguma vantagem estratégica (Dunham, 2002).

Como exemplo de uso desta ferramenta, tem-se o caso de uso da rede de supermercados Walmart, descrito por Serra (2002). Ao analisar os dados de compras dos seus clientes observou-se que em 60% das compras de bonecas Barbie os clientes também compravam barras de chocolate. Com esta informação, decidiu-se rearranjar as prateleiras de modo que os chocolates ficassem mais perto das bonecas. Como resultado, ocorreu um salto nas vendas de chocolates.

As técnicas de mineração de dados podem ser classificadas como descritivas e preditivas. Esta classificação leva em conta a natureza do resultado final gerado pela aplicação de cada técnica (Han e Kamber, 2006).

As técnicas descritivas têm por função determinar características comuns de um conjunto de dados, apresentando padrões e relações existentes nestes dados. Exemplos de técnicas desta classe são (Chakrabarti *et al*, 2008):

- clusterização, na qual são gerados subconjuntos a partir de um conjunto de dados, onde os elementos de cada subconjunto são similares entre si e diferentes dos elementos de outros subconjuntos, como a similaridade baseada em uma função de distância;
- mineração de regras de associação, onde são buscados subconjuntos de dados com alta ocorrência dentro de um conjunto de dados e que apontem para algum padrão associativo, como no exemplo da rede supermercados Walmart, onde a compra da boneca Barbie estava associada à compra de barras de chocolate.

No caso das técnicas preditivas, o objetivo é realizar previsões sobre dados futuros, realizado através de inferências utilizando os dados existentes. Como exemplo de técnicas desta classe, pode-se citar (Han e Kamber, 2006):

- classificação, cujo objetivo é, partindo de um conjunto de dados com classes definidas, estabelecer um modelo ou função capaz de descrever e distinguir essas classes, tornando possível classificar dados cuja classe é desconhecida;
- análise por regressão, que tem funcionamento similar à de classificação, porém, gera como resultado um valor contínuo, ao invés de um valor classificatório.

### **3.3 Criação de Ambientes de BI**

A criação de ambientes de BI (*Business Intelligence*) é uma tarefa desafiadora, visto os diferentes tipos de tecnologias devem ser integrados para compor este ambiente. Buscando facilitar e promover a discussão sobre a criação de ambientes de BI, diversas propostas e estudos de caso foram publicados.

No trabalho de Hong, Zai-wen e Hai-yang (2008), um ambiente de BI é criado para uma rede de supermercados através da construção de um DW utilizando as ferramentas do Microsoft Analysis Services. Neste ambiente são utilizadas técnicas de clusterização de dados para definir perfis de compra dos seus clientes e os tipos de propaganda que mais influenciavam suas compras. O artigo não menciona como os dados do BI serão disponibilizados para os usuários finais.

Em outro trabalho, Xu, Zhang e Jiang (2005) apresentam a elaboração de um ambiente de BI para uma empresa de seguro de vida. Neste ambiente foi adotado um DW para o armazenamento dos dados. Entretanto, o artigo não apresenta a forma como foi conduzida a construção do ambiente, tampouco as ferramentas utilizadas para sua realização.

A construção de um ambiente de BI para um provedor de *Internet* de Taiwan é apresentada no trabalho de Li, Shue e Lee (2008). Neste ambiente, os dados de tráfego de *Internet* dos usuários passam por um processo de ETL e são armazenados em um



DW, utilizando ferramentas do MS-SQL 2000. Sobre estes dados são aplicadas técnicas de clusterização, com o objetivo de definir perfis de uso de *Internet* dos clientes. Estes dados são disponibilizados aos usuários através de um *web site* construído com tecnologia Java e Flash. A construção deste ambiente foi conduzida utilizando uma metodologia própria dos autores.

Tratando especificamente da construção de um DW têm-se os trabalhos de Sahama e Croll (2007) e Wah e Sim (2009) que tratam da construção de um DW para a área de saúde. Sahama e Croll utilizam a ferramenta SAS Data Warehouse Administrator para construção do DW. Já Wah e Sim constroem o DW seguindo uma metodologia proposta, sendo o DW disponibilizado em um banco de dados MySQL. Rifaie *et al* (2008) discutem a construção de um DW para um Sistema de Informação Geográfica (do inglês *Geographycal Information Systems - GIS*), demonstrando os desafios que este tipo de sistema apresenta. Como todos estes trabalhos focam especificamente na construção do DW, nenhum discute como os dados serão disponibilizados aos usuários.

## 4 Ambiente de *Business Intelligence* para e-Gov

Neste capítulo é feita a revisão bibliográfica dos trabalhos relacionados à criação de ambientes de BI (*Business Intelligence*) para e-Gov (Governo Eletrônico), sendo discutidos os contextos de e-Gov em que estes ambientes podem ser aplicados e desafios a serem vencidos para sua criação. A seguir, é apresentada a proposta deste trabalho para criação de um ambiente de BI para e-Gov.

### 4.1 Criação de Ambientes de BI para e-Gov

Trabalhos na literatura tratam a questão da criação dos ambientes de BI para e-Gov. Ruschel (2008) discute a criação de um ambiente de BI para o sistema judiciário. Nele, é escolhido um DW como o local para o armazenamento dos dados gerados pelo processo de ETL. Este DW é construído utilizando a metodologia proposta por Kimball (1996). No entanto, não é apresentada a arquitetura do ambiente gerado, tampouco as ferramentas usadas para sua construção. Outro fator negativo é o trabalho não apresentar como foram executados os passos da metodologia adotada.

Mussi *et al* (2004) apresenta a implantação do DW na Agência Nacional de Vigilância Sanitária (ANVISA). Para tal, foi adotada a metodologia proposta por Kimball (1996), utilizando o banco de dados Oracle para armazenamento dos dados do DW e a ferramenta OLAP MicroStrategy para disponibilização dos dados.

Embora o trabalho de Ruschel (2008) e Mussi *et al* (2004) estejam aplicados no contexto de e-Gov, nenhum analisa como estes ambientes podem ser implementados, considerando as dificuldades particulares de aplicações de e-Gov apresentadas no capítulo 2.1. Além disto, não é feita a discussão teórica sobre quais classes de sistemas de e-Gov se encaixam os ambientes de BI.

## **4.2 Aplicabilidade dos Ambientes de BI para e-Gov**

Como descrito anteriormente, os ambientes de BI tem como principal função disponibilizar em tempo hábil informações que auxiliem os tomadores de decisão no processo de tomada de decisão.

Baseado na arquitetura das aplicações de e-Gov apresentada no capítulo 2.2, observa-se que os ambientes de e-Gov pertencem a camada de e-Business. Nela, estes ambientes podem ser empregados para integrar os dados dos diferentes órgãos e departamentos, permitindo a análise mais completa destes dados.

Utilizando a classificação dos sistemas de e-Gov proposta por Lee, Tan e Trimi (2005), apresentada no capítulo 2.3, pode-se observar que estes ambientes podem ser empregados como aplicações de IEE (Eficiência e Efetividade Interna), sendo utilizados para melhoria na gestão dos recursos e processos internos das organizações. Os ambientes também podem ser utilizados como aplicações de G2G (Governo para Governo), servindo como ferramenta facilitadora da troca de informações entre diferentes órgãos governamentais.

A utilização dos ambientes de BI como aplicações de G2B (Governo para Empresas) e G2C (Governo para o Cidadão) pode trazer grandes benefícios para a comunicação entre o governo e as empresas/cidadãos. No entanto é importante que neste contexto os ambientes considerem que os usuários podem possuir pouco ou nenhum conhecimento sobre a forma como as ferramentas tradicionais de BI funcionam, sendo assim necessária a criação de aplicações mais simples que permitam a estes acessar as informações.

Por fim, a aplicação destes ambientes também se encaixa na classe de Infraestrutura Global, auxiliando no acompanhamento dos atributos de qualidade dos serviços providos pelo e-Gov.

Para a utilização dos ambientes de BI, vemos como uma das principais barreiras os baixos orçamentos dos setores de TI (Tecnologia da Informação), resultando na

perda de pessoal qualificado e poucos investimentos para treinamento e aquisição de novos aplicativos, como discutido no capítulo 2.1.

Frente ao exposto, este trabalho propõe a utilização de tecnologias de código aberto (do inglês *open source*), que possuam manuais e documentos gratuitos para auxiliar da instalação e utilização das mesmas, além de canais de comunicação, como fóruns de discussão *online*, que possam auxiliar na resolução de dúvidas e problemas. Desta forma, é possível que sejam criados ambientes de BI sem nenhum investimento na aquisição de licenças de *software* e que os responsáveis pela criação do ambiente possam aprender a utilizar as tecnologias e/ou sanar problemas ou dúvidas através dos recursos disponíveis.

### **4.3 Ambiente Proposto**

Para a aplicação dos ambientes de BI para e-Gov, este trabalho propõe um ambiente composto pela integração de diferentes ferramentas e tecnologias de código aberto, cobrindo desde a obtenção dos dados utilizando as técnicas de ETL (Extração, Transformação e Carga) à disponibilização de ferramentas que permitam aos usuários finais a análise e manipulação destes dados.

O ambiente de BI proposto utiliza como arquitetura de referência a arquitetura definida por Moss e Atre (2003). Nesta arquitetura, ilustrada na Figura 4.1a, o ambiente é dividido em três camadas: a Camada de ETL, a Camada de Armazenamento e Disponibilização de Visões dos Dados e a Camada de Aplicações para os Usuários Finais. A Figura 4.1b apresenta o ambiente proposto e as tecnologias que o compõe.

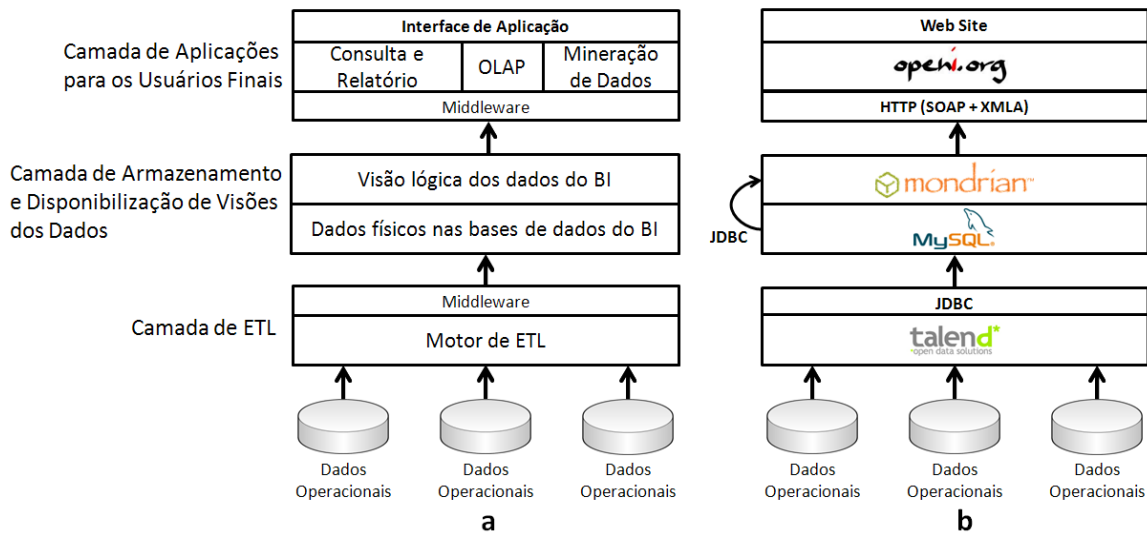


Figura 4.1 – a) Arquitetura proposta por Moss e Atre (2003) b) Ambiente de BI proposto.

A seguir serão descritas as ferramentas e tecnologias de cada camada, mostrando suas principais funcionalidades, a integração entre cada componente da arquitetura e as razões que motivaram a escolha de cada ferramenta.

### 4.3.1 Camada de ETL

A camada de ETL é responsável pelo processo de ETL, no qual os dados dos repositórios de dados operacionais são extraídos, transformados e armazenados nas bases de dados do ambiente de BI. Nesta arquitetura, a camada é dividida em duas subcamadas: *Motor ETL*, onde o processo de ETL é executado e *Middleware*, responsável por mediar a comunicação entre a subcamada *Motor ETL* e a subcamada *Dados Físicos nas Bases de Dados do BI* (Moss e Atre, 2003).

Para implementação da subcamada *Motor ETL*, foi adotada a ferramenta Talend Open Studio (TOS) (Talend, 2010), a qual é especializada na integração e migração de dados.

Nela, é possível a leitura e escrita de dados armazenados em diversos formatos, como Excel, CSV e XML. Também permite conexão a SGBD relacionais, hierárquicos e multidimensionais. Aos dados lidos podem-se aplicar diferentes tipos de

transformações, como a normalização de campos, cruzamento de dados utilizando uma chave de identificação e tratamentos em campos texto.

O processo de ETL é concebido na ferramenta por meio da composição e configuração de componentes. Tipicamente, cada componente é responsável por uma etapa do processo, como a leitura do dado de uma tabela ou a aplicação de uma transformação sobre um dado lido. A integração dos componentes é denominada *job*.

Um exemplo simples desse processo é apresentado na Figura 4.2, onde é apresentado um *job* composto por quatro componentes. Os componentes **a** e **b** são responsáveis pela leitura de dados das bases de dados operacionais (extração). Os dados lidos são combinados (transformação) pelo componente **c** e armazenados (carga) na base de dados destino pelo componente **d**.

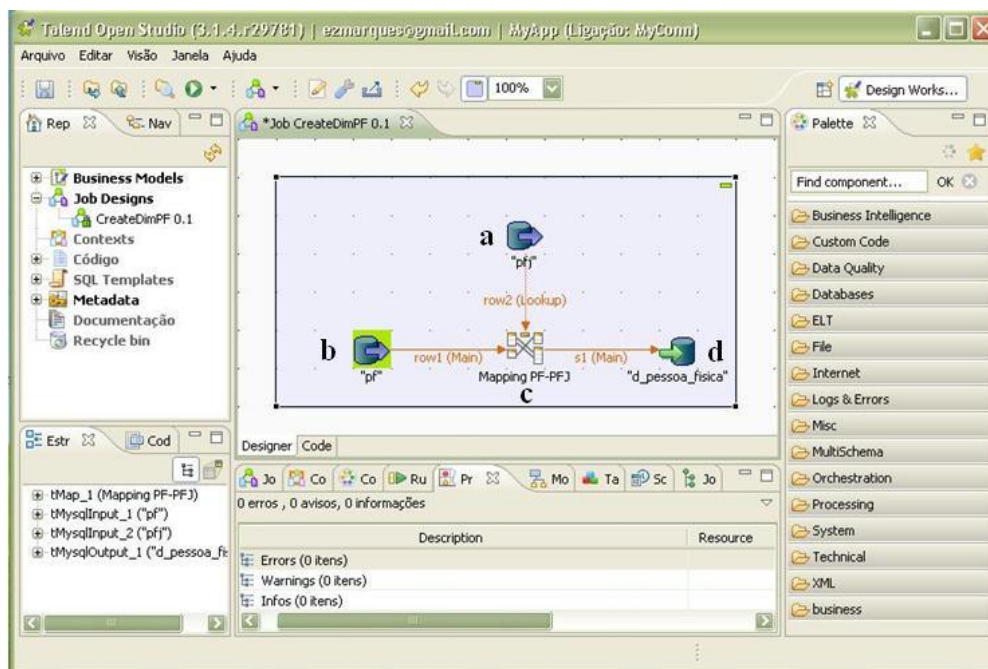


Figura 4.2 - Exemplo de utilização da ferramenta Talend Open Studio.

Outra opção que poderia ser adotada para a criação da subcamada *Motor ETL* é a ferramenta Pentaho Data Integration (PDI), também conhecida por Kettle (Kettle, 2010). Nesta ferramenta o processo de ETL é elaborado de maneira similar a ferramenta TOS. A escolha pelo TOS levou em conta os critérios (Tabela 4.1):

- Documentação: a documentação completa do TOS está disponível para acesso na página da ferramenta, enquanto a documentação do PDI apenas apresenta uma rápida introdução à ferramenta, indicando como referência complementar livros que não possuem livre acesso;
- Execução: o TOS permite que o processo gerado seja exportado, gerando arquivos *jar* e um arquivo *bat* e *sh*, possibilitando a execução do processo de ETL em máquinas que não possuem a ferramenta instalada. O PDI não possui recursos para que o processo de ETL seja executado em máquinas que não possuam a ferramenta instalada;
- Desempenho: este critério não pode ser levado em consideração na escolha da ferramenta por não ter-se conseguido encontrar um comparativo (*benchmark*) de uma fonte idônea de livre acesso.

Tratando da subcamada *Middleware*, optou-se pelo padrão JDBC para mediar a comunicação. Este padrão, amplamente adotado pela comunidade de desenvolvedores Java, permite que a ferramenta Talend Open Studio se comunique com o SGBD MySQL, solução adotada para o armazenamento dos dados no ambiente, descrita a seguir.

Tabela 4.1 - Comparativo entre as soluções para camada de ETL.

Característica	Talend Open Studio	Pentaho Data Integration
Documentação	Completa disponível no site	Introdutória disponível no site
Execução	Exportação do processo para execução fora da ferramenta	Apenas execução na ferramenta
Desempenho	Não foi avaliado por não ser encontrado benchmark de fonte idônea aberto à comunidade	

### 4.3.2 Camada de Armazenamento e Disponibilização de Visões dos Dados

A camada de Armazenamento e Disponibilização dos Dados é responsável pela gestão dos dados utilizados no ambiente de BI, resultantes do processo de ETL

executado pela camada anteriormente descrita. Para tanto, esta camada é dividida nas subcamadas *Dados físicos nas bases de dados do BI*, com função de prover mecanismos para o armazenamento dos dados, e *Visão lógica dos dados do BI*, responsável por gerar representações claras dos dados para as camadas superiores (Moss e Atre, 2003).

Para o armazenamento dos dados na subcamada *Dados físicos nas bases de dados do BI* foi escolhido o SGBD relacional MySQL (MySQL, 2010). A escolha deste SGBD deve-se ao suporte a diversos tipos de índices (*B-Tree* e *Hash*, por exemplo) e rapidez na carga e acesso aos dados, através do uso de tabelas MyISAM, tecnologia oferecida pelo MySQL.

Como realização da subcamada *Visão lógica dos dados do BI*, foi adotado o servidor OLAP Mondrian (Mondrian, 2010). O Mondrian é um servidor OLAP (Processamento Analítico de Tempo Real) escrito em Java que permite a execução de consultas multidimensionais, no formato *Multi Dimensional Expressions* (MDX), em uma base de dados relacional. Para isso, é necessária a criação um arquivo XML de mapeamento multidimensional dos dados relacionais, definindo-se cubos de dados com seus fatos e dimensões. Para auxiliar nesta tarefa, juntamente com o servidor é disponibilizada a ferramenta Mondrian Schema Workbench (Figura 4.3), que provê interface com facilitadores para confecção do XML de mapeamento.

O formato MDX é um padrão definido pela Microsoft para a execução de consultas sobre cubos de dados, possibilitando o acesso a dados multidimensionais de maneira mais simples e intuitiva. De certa forma, é possível afirmar que o MDX é para as bases de dados multidimensionais o que o SQL é para as bases de dados relacionais, salvo que o MDX retorna em suas consultas cubos de dados, enquanto o SQL retorna tabelas. Além disto, o SQL possui funções de definição de dados (DDL), que atualmente não estão presentes na especificação do MDX. Além do Mondrian, diversas aplicações suportam consultas MDX, como MicroStrategy, SAS e Oracle Essbase (Xmla, 2010).



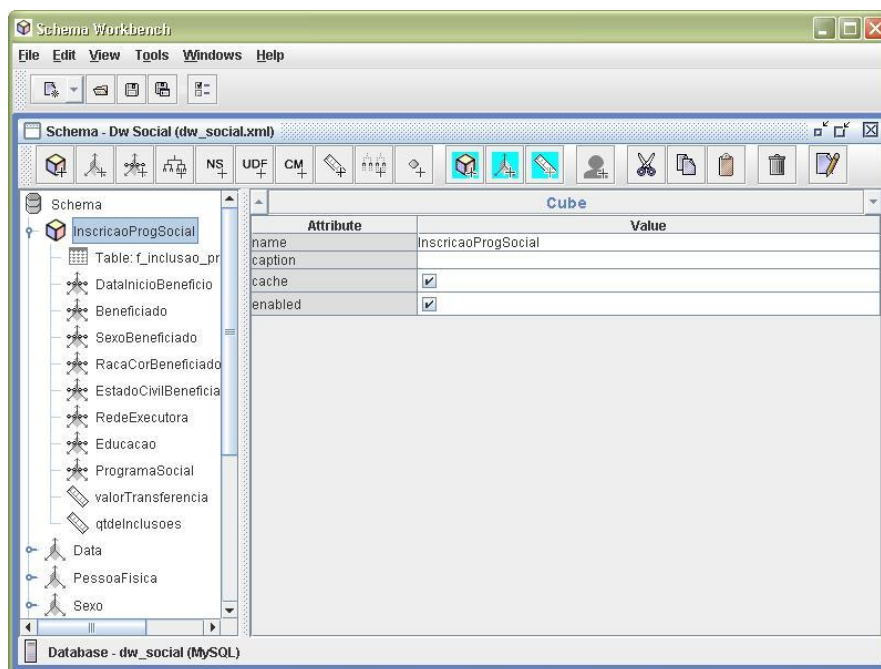


Figura 4.3 - Ferramenta Mondrian Schema Workbench.

A criação de visões dos dados traz grandes benefícios ao ambiente de BI uma vez que, por meio dela, é possível definir diferentes visões dos dados para diferentes usuários, além de possibilitar a apresentação mais clara dos dados do DW para o usuário final.

Pode-se também implementar esta camada pelo servidor Palo (Palo, 2010). Entretanto, optou-se no trabalho pelo servidor Mondrian, por este implementar o padrão *XML for Analysis* (XMLA) para comunicação na versão de uso livre, enquanto este mesmo padrão só é suportado na versão paga do servidor Palo. Outro benefício do Mondrian é a possibilidade de gerenciar o controle de acesso aos dados expostos pelo servidor (Tabela 4.2). Tanto o protocolo XMLA quanto o controle de acesso aos dados serão discutidos posteriormente no trabalho.

Tabela 4.2 - Comparativo entre as soluções para camada de Visão Lógica dos Dados.

Característica	Mondrian	Palo
Documentação	XMLA	XMLA apenas na versão paga
Execução	Sim	Não

### 4.3.3 Camada de Aplicações para os Usuários Finais

A camada de Aplicações para os Usuários Finais tem como função disponibilizar soluções que permitam aos usuários do ambiente de BI acessar e analisar os dados contidos no ambiente, além de definir as formas como as visões dos dados serão acessadas e como as aplicações serão disponibilizadas aos usuários (Moss e Atre, 2003).

A aplicação OpenI (OpenI, 2010) foi selecionada como ferramenta provida aos usuários do ambiente proposto. O OpenI é um aplicativo *web* desenvolvido em Java que permite a execução de consultas sobre bases de dados multidimensionais. Nele, o usuário seleciona um cubo de dados e, através da interface gráfica, pode navegar nos dados utilizando as principais funções de um cliente OLAP. Também é possível que o usuário navegue pelos dados utilizando a linguagem MDX para consultas mais elaboradas. O resultado pode ser visualizado na forma de gráficos (Figura 4.4) ou de tabelas multidimensionais (Figura 4.5). Além disto, estas consultas podem ser salvas e, posteriormente, visualizadas e editadas.

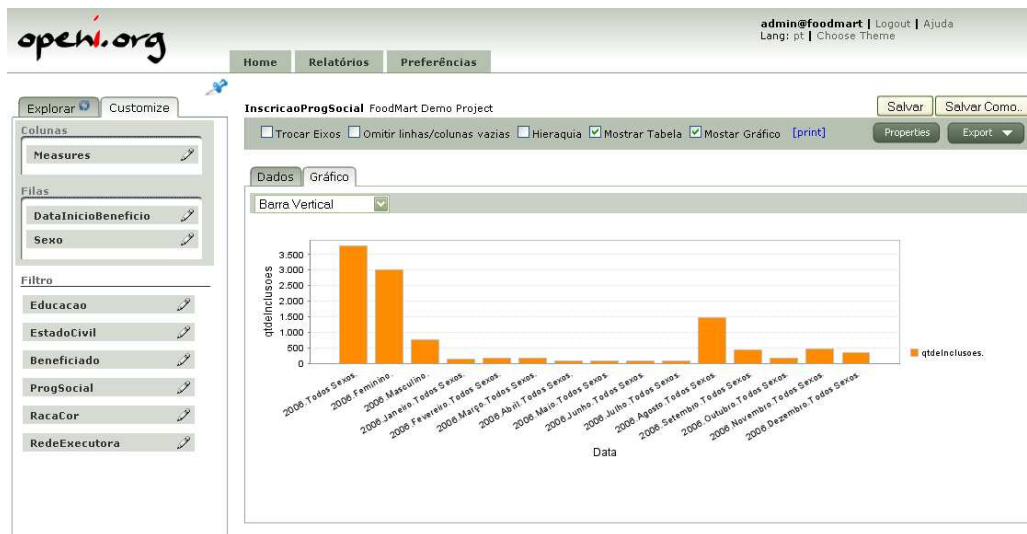


Figura 4.4 - Resultado de uma consulta no OpenI na forma de gráfico de barras.

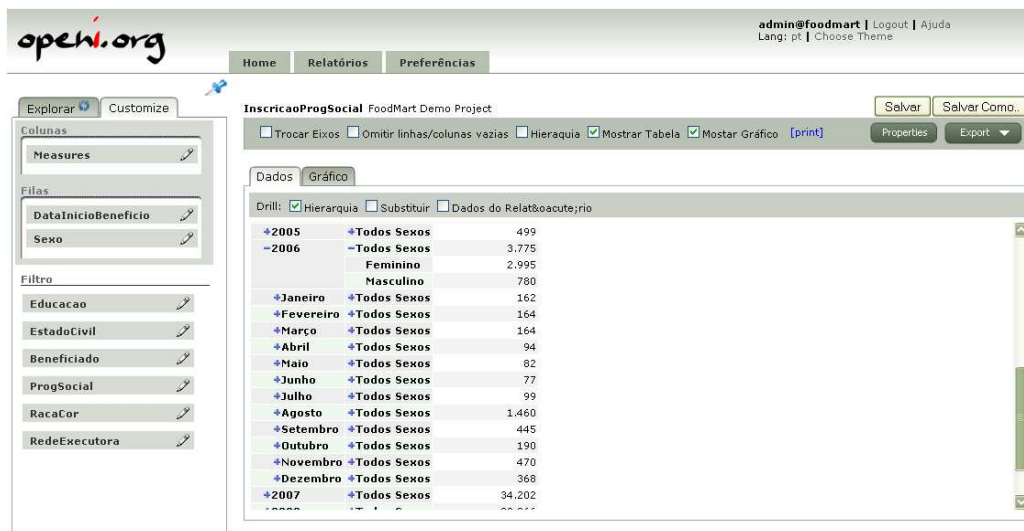


Figura 4.5 - Resultado de uma consulta na OpenI na forma de tabela multidimensional.

Outra funcionalidade importante do OpenI é a criação e disponibilização de painéis de indicadores (do inglês *dashboards*), onde consultas previamente salvas são agrupadas para a rápida visualização de dados sensíveis de negócio. Para cada consulta pode ser escolhida uma diferente forma de apresentação dos dados, semelhante ao que ocorre na visualização de resultados de consultas.

Além do OpenI, podemos encontrar outros clientes OLAP de uso livre, como o Palo Web (Palo, 2010), JPivot (JPivot, 2010) e JasperAnalysis (JasperSoft, 2010). No entanto, nenhum destes clientes permite que o usuário navegue pelos dados de um cubo de dados através de interface de arrastar e soltar (do inglês *drag & drop*), fator decisivo na escolha pelo OpenI. No caso do Palo Web e do JasperAnalysis este recurso existe, mas está apenas disponível nas versões pagas (Tabela 4.3).

Tabela 4.3 - Comparativo entre as soluções de clientes OLAP.

Característica	OpenI	Palo Web	JPivot	Jasper Analysis
Interface drag & drop para navegação no cubo de dados	Sim	Apenas na versão paga	Não	Apenas na versão paga

A comunicação entre o OpenI e o servidor OLAP Mondrian é feita via HTTP, usando os protocolos *Simple Object Access Protocol* (SOAP) e *XML for Analysis* (XMLA).

O protocolo SOAP tem como objetivo estruturar a troca de informações através do HTTP. Utilizando o padrão XML, este protocolo é composto por um envelope que possui duas *tags*: **header**, onde estão alocadas as informações sobre como a mensagem deve ser processada e **body**, que contém as informações da mensagem transmitida (Snell, Tidwell e Kulchenko, 2002).

Utilizando o SOAP, o protocolo XMLA define um conjunto de mensagens XML para a comunicação entre uma aplicação cliente e um servidor OLAP, estabelecendo um canal de comunicação cliente-servidor interoperável entre diversas ferramentas de diferentes empresas. Os tipos de mensagens definidos pelo protocolo são (Xmla, 2010):

- *Discover*: utilizada para consultar propriedades sobre o servidor OLAP, como a lista de fontes de dados disponíveis;
- *Execute*: utilizada para executar consultas em uma fonte de dados específica, usando a linguagem MDX.

#### 4.3.4 Aspectos de Segurança do Ambiente

Um dos principais atributos de qualidade de todo ambiente de *software* é a segurança dos dados. Para a criação de um ambiente onde haja segurança dos dados, é necessário considerar diversas questões, como a configuração de *firewall* nas máquinas expostas na *Internet*, comunicação segura dos dados utilizando algoritmos criptográficos, controle de acesso aos dados, políticas de atualização de senhas e políticas de fortalecimento de senhas dos usuários. Neste tópico, serão discutidos os aspectos de segurança na comunicação de dados e no acesso aos dados do ambiente proposto.

#### 4.3.4.1 Segurança na comunicação de dados

No ambiente proposto, existem quatro pontos de comunicação de dados. Três destes são responsáveis pela comunicação entre as camadas e o último é utilizado para os clientes acessarem as aplicações, feita pela *Internet* através de um *web site* (Figura 4.6). Nestes pontos, a comunicação é feita utilizando apenas dois meios:

1. comunicação através do *middleware* JDBC, na subcamada *Middleware* da camada de ETL e na comunicação entre as subcamadas da camada de Armazenamento e Disponibilização dos Dados;
2. comunicação através do protocolo HTTP, na subcamada *Middleware* da camada de Aplicações para os Usuários Finais e para disponibilização do OpenI para o usuário final.

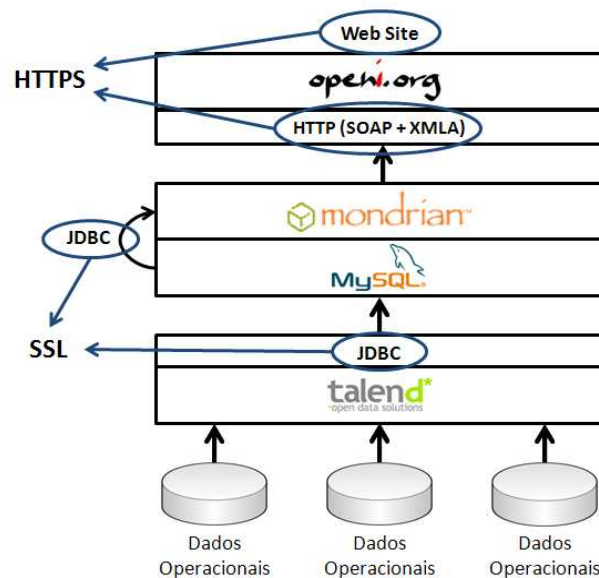


Figura 4.6 - Soluções adotadas para a segurança da comunicação de dados.

No caso da comunicação através do JDBC, a segurança é responsabilidade do fornecedor do *middleware*. No cenário proposto, são utilizadas implementações do JDBC para comunicação com os SGDB MySQL. Neste ambiente (MySQL, 2010), a implementação do JDBC utilizada é fornecida pela mesma empresa que desenvolve do

SGDB e a segurança é garantida utilizando o protocolo *Secure Socket Layer* (SSL) (Freier, Karlton e Kocher, 1996).

No caso dos pontos de comunicação HTTP, a segurança pode ser garantida através da mudança para o protocolo *Hypertext Transfer Protocol Secure* (HTTPS) que, também utilizando o protocolo SSL, ou mais recentemente o protocolo *Transport Layer Security* (TLS) (Dierks e Rescorla, 2008), garante a segurança dos dados transmitidos (Rescorla, 2000).

#### **4.3.4.2 Segurança no acesso aos dados**

No ambiente proposto nota-se que o controle no acesso aos dados é uma preocupação da camada de Armazenamento e Disponibilização dos Dados, especificamente na subcamada de Disponibilização dos Dados, uma vez que a responsabilidade desta é prover as visões dos dados para as aplicações dos usuários finais.

No Mondrian, a segurança no acesso aos dados é feita através da técnica de Controle de Acesso Baseados em Papéis (do inglês *Role-Based Access Control* – RBAC) (Mondrian, 2010).

A técnica RBAC visa simplificar a implementação e gerência do controle de acesso em sistemas de informação. Nesta técnica, o usuário adquire privilégios de execução de funcionalidades no sistema e/ou acesso aos dados por meio da atribuição de papéis ao mesmo. A cada papel do sistema é associado um conjunto de permissões. Desta maneira, caso haja mudanças nas permissões de um papel, seja pela aquisição ou perda de uma permissão, todos os usuários que o possuem serão afetados (Sandhu, Ferraiolo e Kuhn, 2000).

Utilizando os recursos para definição de perfis no Mondrian, é possível configurar permissões de acesso aos dados nos mais diversos níveis de granularidade. Um papel pode possuir a permissão de acesso a um *data mart*, a um cubo de dados e/ou a uma hierarquia toda ou a parte dela. Também é possível fazer o controle de registros específicos do cubo de dados como, por exemplo, proibir que os dados sobre

as compras de um produto em uma loja numa data específica sejam visualizados por papéis específicos.

## **5 Estudo de Caso**

Este capítulo descreve a aplicação do ambiente proposto para a construção de um ambiente BI para análise dos dados operacionais da Assistência Social da Prefeitura Municipal de Campinas, SP. A aplicação da proposta busca analisar a viabilidade do ambiente de BI em um cenário real de e-Gov, permitindo observar características deste ambiente.

### **5.1 Contexto de Aplicação da Proposta**

A área de Assistência Social da prefeitura tem como missão “resgatar a cidadania da população que se encontra em situação de vulnerabilidade social causada pela pobreza e exclusão”. Para isso são desenvolvidos diversos programas sociais, buscando prover aos cidadãos novas possibilidades que torne possível sua emancipação financeira. Estes programas sociais abrangem diversas áreas, tais como disponibilização de microcrédito, ações de inclusão digital e cursos de capacitação (Prefeitura Municipal de Campinas, 2010).

Para o gerenciamento dos dados operacionais desta área, foi adotado o módulo de Gestão Social do Sistema Integrado de Governança Municipal (SIGM). O SIGM é um ERP voltado às necessidades de gestão das prefeituras. Desta forma, são oferecidos mecanismos para o gerenciamento de todos os serviços, registros de cidadãos, gerência de processos e dados relevantes para a administração do município. Este sistema é desenvolvido sobre uma arquitetura de múltiplas camadas, utilizando a tecnologia EJB para a distribuição dos objetos de negócio e SGBD relacional para o armazenamento dos dados (Tilli *et al*, 2008).

No módulo de Gestão Social, podem ser armazenados e manipulados todos os dados relacionados aos programas sociais executados pelo governo municipal, aos



cidadãos que participam destes programas e às redes executoras responsáveis pela realização das ações proporcionadas por esses programas.

Neste cenário, o ambiente de BI foi construído através de duas iterações da metodologia descrita a seguir.

## **5.2 Metodologia Adotada**

Para guiar a execução do estudo de caso, foi definida uma metodologia iterativa e incremental com cinco passos sequenciais, baseada na metodologia proposta por Kimball *et al* (1998):

- i. definição dos cubos de dados candidatos a compor o DW do ambiente e seleção de um conjunto de cubos para ser implementado;
- ii. análise conceitual, lógica e física dos cubos de dados selecionados;
- iii. configuração do processo de ETL para preenchimento do DW com os dados provenientes dos repositórios de dados operacionais;
- iv. criação das visões de dados dos cubos selecionados;
- v. configuração da aplicação disponibilizada aos usuários finais para acesso às visões de dados.

No **passo i**, foi adotada a análise orientada em dados (Inmon, 2005). Neste tipo de análise, as informações utilizadas como entrada para a geração do DW são somente os modelos das bases de dados operacionais que serão a fonte de dados para povoar o ambiente de BI.

No **passo ii**, a modelagem conceitual foi executada buscando definir a estrutura conceitual dos cubos selecionados, sem se preocupar com a forma que os dados seriam

armazenados e acessados. Essa preocupação, no entanto, fez parte da modelagem lógica e física dos cubos de dados.

Para documentação dos cubos de dados candidatos no **passo i** e documentação do modelo conceitual do **passo ii** foram utilizados modelos da linguagem UML com o perfil de modelagem de dados multidimensionais proposto por Luján-Mora, Trujillo e Song (2006).

Este perfil possui três níveis de detalhamento das estruturas do DW. No primeiro, são mostradas apenas as estrelas que compõe o DW e as dependências entre elas. No segundo, são mostrados os fatos e as dimensões relacionadas a cada um. No terceiro, é descrita a estrutura plena de um fato e das dimensões relacionadas a este, sendo definidas as métricas dos fatos, os atributos e os níveis hierárquicos das dimensões.

Ainda no **passo ii**, para a modelagem lógica e física dos cubos foi utilizado o modelo estrela descrito no capítulo 3.2.2.

No **passo iii** foram feitas as configurações na ferramenta Talend Open Studio para acesso às bases de dados operacionais, limpeza e ajustes dos dados operacionais e preenchimento dos cubos de dados do DW.

No **passo iv**, as visões dos dados armazenadas no DW foram criadas provendo representações claras dos dados aos usuários finais. A estrutura das visões de dados gerada é similar à estrutura do modelo conceitual dos cubos de dados. Isto se justifica devido ao fato do modelo conceitual organizar os dados da forma mais próxima como o usuário final compreende os dados, desconsiderando questões de desempenho de consultas ou características específicas da tecnologia de armazenamento de dados do ambiente (Rizzi *et al*, 2006).

Por fim, no **passo v** foi feita a configuração do OpenI para acessar as visões de dados. Esta configuração se dá através da interface de administração do sistema.

### 5.3 Primeira Iteração

A primeira iteração da metodologia teve por intuito testar o ambiente proposto em um cenário real, buscando identificar possíveis deficiências que não haviam sido observadas durante os testes experimentais do ambiente. Com essa missão, decidiu-se executar uma iteração com poucos dados, mas que fossem suficientes para mostrar o potencial da solução.

No passo i, os modelos ER das bases de dados do módulo de assistência social do SIGM e dos módulos relacionados foram utilizados para o levantamento dos cubos candidatos. Estes modelos possuem aproximadamente cem tabelas. Como resultado deste passo, foram levantados três cubos de dados candidatos (Figura 5.1) e apenas o cubo de dados *Inclusão Programa Social* foi selecionado para implementação. Este cubo de dados representa o evento de uma pessoa ser incluída em um programa social.

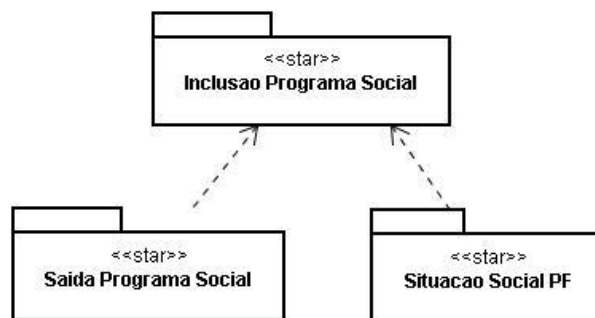


Figura 5.1 - Diagrama dos cubos de dados candidatos da primeira iteração.

Terminado o passo i, iniciou-se a modelagem conceitual, lógica e física do cubo de dados selecionado. O resultado da modelagem conceitual é apresentado na Figura 5.2 e da modelagem lógica e física na Figura 5.3.

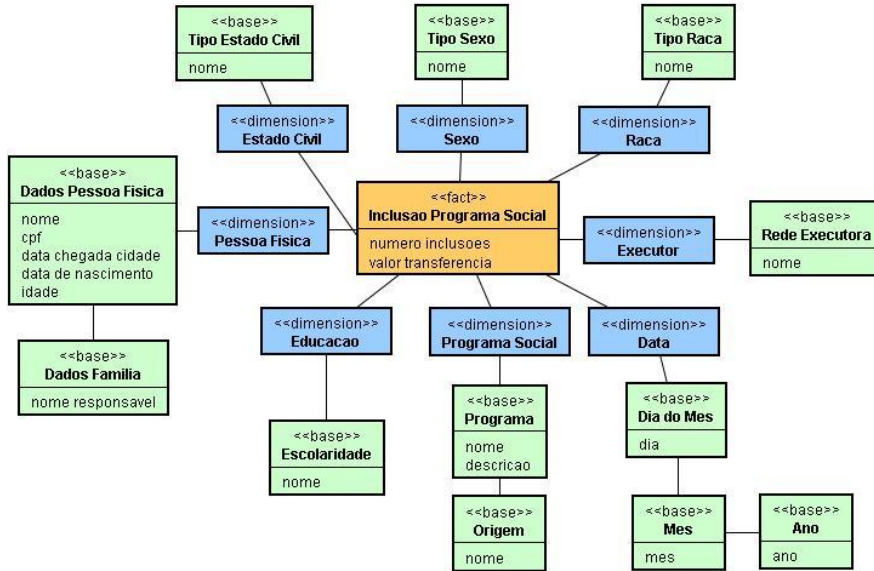


Figura 5.2 - Diagrama conceitual do cubo de dados *Inclusão Programa Social*.

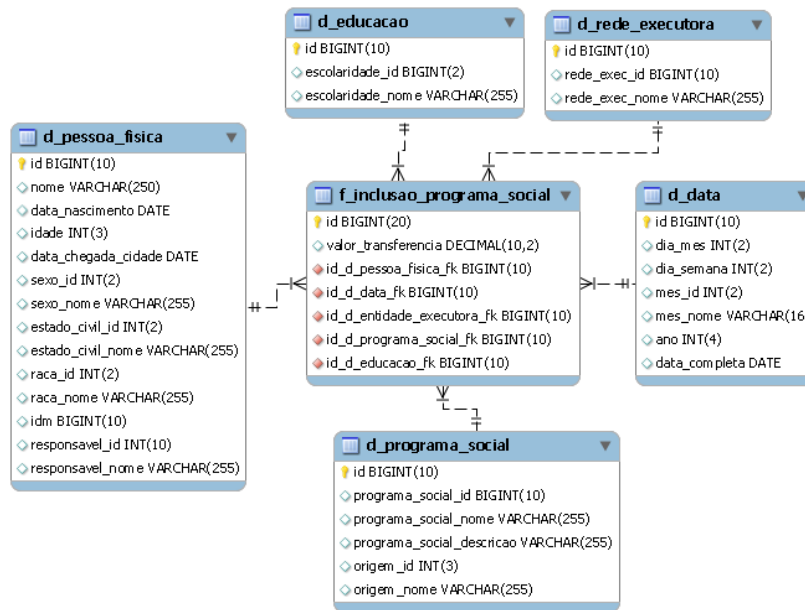


Figura 5.3 - Diagrama físico do cubo de dados *Inclusão Programa Social*.

Concluída a modelagem do cubo de dados, iniciou-se a configuração da ferramenta Talend para a execução do ETL nos dados operacionais. Nesta etapa foram criados *jobs* particulares para povoar cada dimensão e o fato. Por fim, foi criado um *job* para orquestrar a execução de todos os *jobs* criados, tornando a execução completa do processo de ETL mais simples.

Paralela à construção do processo ETL, a visão dos dados do DW foi configurada. Como dito anteriormente, a visão de dados provida ao usuário possui estrutura similar ao modelo conceitual do cubo de dados.

Concluindo a primeira iteração, foi executado o passo v, onde a aplicação OpenI foi configurada para o acesso à visão de dados exposta no OLAP Server Mondrian.

Finalizados os passos da primeira iteração, o usuário já pode ter acesso ao cubo de dados utilizando as interfaces de consulta do OpenI. Embora na primeira iteração tenha sido selecionado apenas um cubo de dados para implementação, o ambiente de BI resultante já possuía informações suficientes para responder a diversas perguntas, como:

- Para cada ano, em quais meses houve mais inclusões em programas sociais (Figura 5.4)?
- Qual o percentual de analfabetos inclusos em um determinado programa social para um determinado período?
- Das pessoas beneficiadas por um determinado programa social, quantas são pessoas nascidas em outros municípios e atualmente residem na cidade?

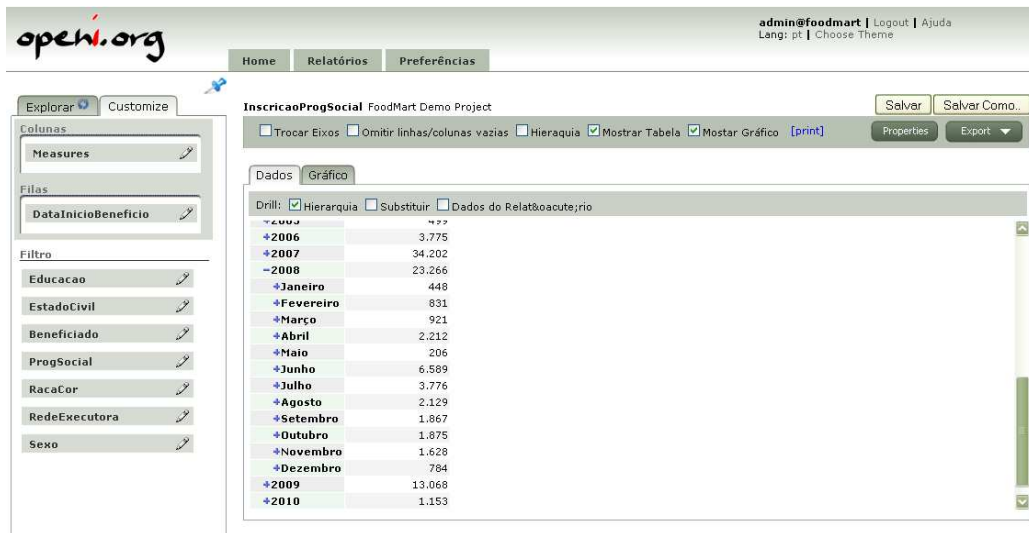


Figura 5.4 - Consulta executada sob dados fictícios no ambiente do estudo de caso.

## 5.4 Segunda Iteração

Na segunda iteração, buscou-se fazer uma análise mais profunda dos dados disponíveis para a criação do DW, melhorando o cubo de dados implementado na iteração anterior e trazendo novos cubos de grande importância de negócio para o ambiente de BI.

Novamente no passo i, foram analisados os modelos ER da base de dados do módulo de assistência social do SIGM e dos módulos relacionados. Desta vez, foram levantados nove cubos candidatos. Destes, foram escolhidos os cubos de dados *Saída Programa Social* e *Atendimento* para serem incluídos no ambiente, além do aprimoramento do cubo de dados *Inclusão Programa Social*. Estes cubos de dados representam, respectivamente, as ações de desvincular uma pessoa de um programa social e prover um atendimento social, como o pagamento de uma bolsa ou entrega de uma cesta básica.

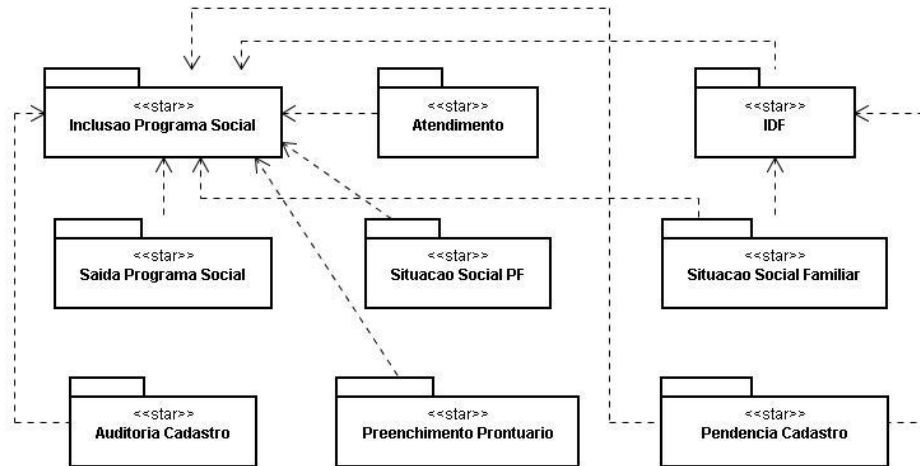


Figura 5.5 - Diagrama dos cubos de dados candidatos da segunda iteração.

Selecionados os cubos de dados, iniciou-se o passo ii, em que foram gerados os modelos conceituais, lógicos e físicos dos novos cubos de dados, além do aprimoramento dos modelos do cubo de dados *Inclusão Programa Social* gerados na iteração anterior. Na Figura 5.6 pode-se observar que foi incluída a dimensão *Localização* no cubo de dados *Inclusão Programa Social* e que os cubos de dados possuem as mesmas dimensões, salvo o cubo de dados *Atendimento*, que possui uma dimensão extra para identificar qual o tipo de atendimento prestado. O modelo conceitual detalhado e o modelo físico do cubo de dados *Atendimento* estão ilustrados na Figura 5.7 e na Figura 5.8.

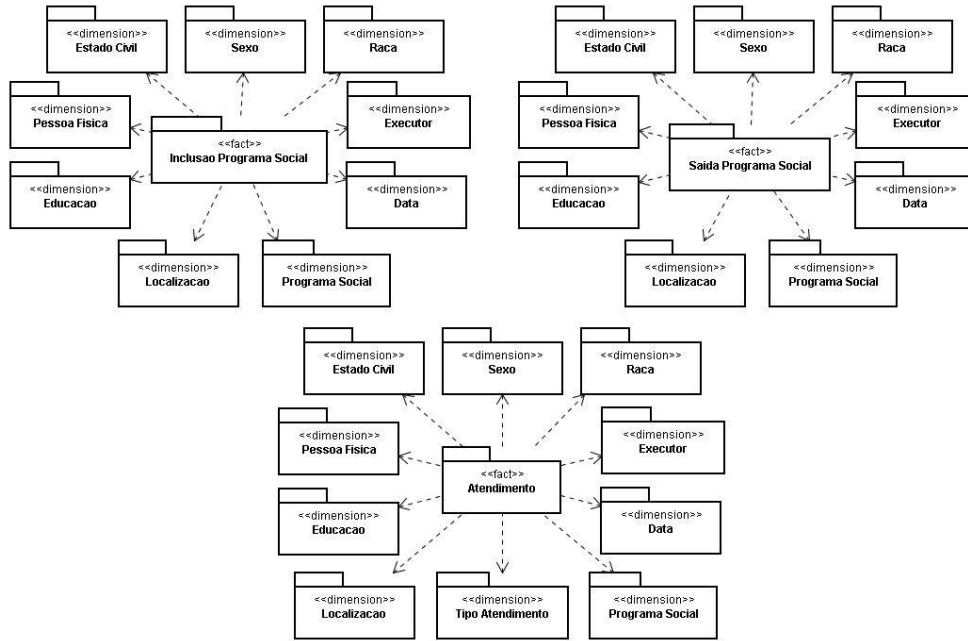


Figura 5.6 - Modelo conceitual dos cubos de dados seleccionados na segunda iteracao.

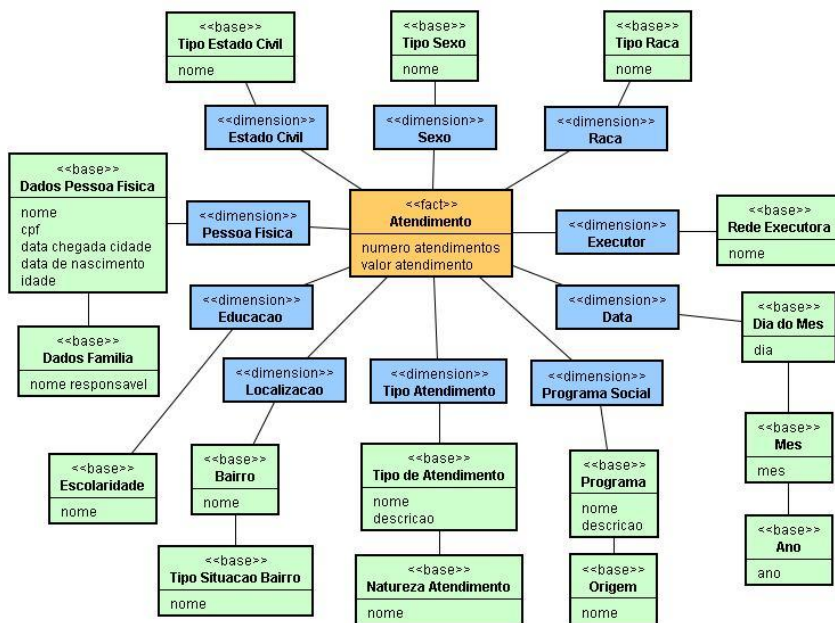


Figura 5.7 - Diagrama conceitual do cubo de dados *Atendimento*.



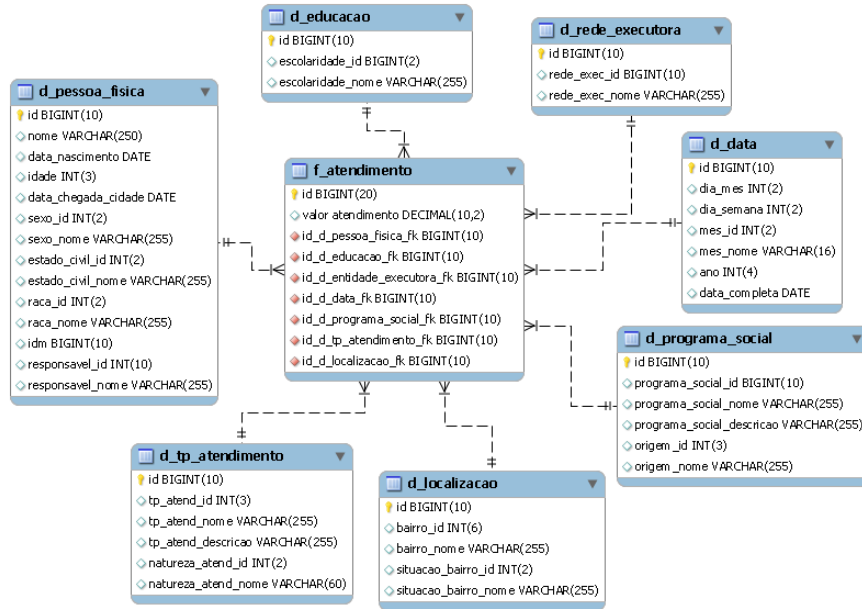


Figura 5.8 - Diagrama físico do cubo de dados *Atendimento*.

A execução dos passos iii e iv foi mais simples quando comparado com a iteração anterior. Isso ocorreu principalmente devido ao reuso dos artefatos gerados na primeira interação. Finalizando a segunda iteração, foi executado o passo v, onde foi configurado o acesso às visões dos novos cubos de dados.

Esta iteração agregou grande poder analítico ao ambiente desenvolvido, pois, além de incluir dois novos cubos de dados, foi adicionada a dimensão *Localização*, a qual permite análises considerando os bairros do município.

## 5.5 Análise da Aplicação do Ambiente de BI

A aplicação do ambiente proposto em um estudo de caso foi de grande relevância para o trabalho, permitindo verificar diversas características sobre este ambiente.

Ao fim de apenas duas iterações da metodologia proposta criou-se um ambiente de análise de dados robusto, onde um usuário é capaz de obter uma visão clara da forma como os programas sociais estão sendo implementados e executados na cidade.

Utilizando o cubo de dados *Inclusão Programa Social* pode-se, por exemplo, acompanhar a eficácia de ações municipais para inclusão dos cidadãos nos programas sociais, além de possibilitar a descoberta de padrões até então desconhecidos, como períodos do mês ou até dias da semana onde ocorre a maior parte das inclusões. Análises similares podem ser feitas para os cubos de dados *Saída Programa Social e Atendimento*, onde é possível extrair informações referentes à eficácia e a aderência dos programas sociais.

A forma incremental como a metodologia foi proposta trouxe flexibilidade para o ambiente aumentar seu potencial progressivamente, permitindo que informações extraídas que não possam ser claramente explicadas com os dados presentes no ambiente venham a ser esclarecidos pela inclusão de novos dados.

O fato da análise dos dados neste ambiente ser feita toda manualmente pode ser considerado como um ponto negativo do ambiente. Este problema pode ser resolvido através da inclusão de novas ferramentas ao usuário final focadas neste tipo de atividade, como ferramentas de Mineração de Dados.

Neste sentido, a utilização de uma arquitetura de referência em camadas trouxe flexibilidade para o ambiente acomodar novas tecnologias, como o caso da adição de uma ferramenta de Mineração de Dados ou da substituição de partes já definidas, como a troca do cliente OLAP OpenI por um cliente mais sofisticado, como o MicroStrategy (MicroStrategy, 2010). Esta flexibilidade é importante para permitir que o ambiente evolua de forma simples, atendendo às necessidades que forem requeridas pelos usuários.

Outro ponto que merece destaque é a questão de viabilidade para construção de ambientes de BI na esfera de e-Gov criada com esta proposta. Mesmo departamentos com orçamento relativamente baixo dentro de uma prefeitura podem criar um ambiente sofisticado de análise de dados com baixo custo utilizando o ambiente proposto. Isso é possível devido ao emprego de tecnologias livres neste ambiente, o que implica em custo zero com licenças de *software*. Entretanto, é importante ressaltar a necessidade

da capacitação do corpo técnico para que os recursos das ferramentas que compõe o ambiente sejam utilizados corretamente.

## 6 Conclusão

Atualmente, um dos principais desafios enfrentados pelas organizações é a necessidade de métodos para a análise dos dados gerados por seus sistemas de informação. Esta análise tem como objetivo extrair informações utilizadas, por exemplo, na melhoria da gestão estratégica dos recursos das organizações e na criação de novas oportunidades de negócio. Ambientes de *Business Intelligence* são soluções amplamente utilizadas para auxiliar na tarefa da análise de dados nas organizações privadas. Entretanto, poucos trabalhos discutem a aplicação dos mesmos nas organizações públicas.

Este trabalho promoveu a discussão sobre como estes ambientes podem auxiliar as organizações públicas na análise dos dados gerados pelos seus sistemas de informação. Também é proposta uma abordagem para a criação de um ambiente de BI para e-Gov baseada em uma arquitetura de referência e composta por ferramentas e tecnologias de código aberto, as quais tiveram suas características gerais, a forma de comunicação e os critérios de escolha demonstrados.

Complementando a abordagem, foi apresentada uma metodologia composta por cinco passos para auxiliar na implantação deste ambiente. A abordagem proposta foi aplicada em um estudo de caso, onde foi construído um ambiente de BI para o Departamento de Assistência Social da cidade de Campinas.

O desenvolvimento do ambiente sobre uma arquitetura de referência trouxe clareza e flexibilidade, tornando a função de cada componente mais clara e facilitando a sua evolução e/ou substituição. A utilização de tecnologias de código aberto foi de grande relevância para a abordagem proposta, por possibilitar a criação de um ambiente de BI onde não existem custos com licenças de *software*. Foi possível encontrar ferramentas de *software* com documentação disponível para acesso gratuito e com uma comunidade ativa trabalhando na evolução destas ferramentas e auxiliando na resolução de dúvidas sobre o uso destas.

Desta forma, podem-se destacar como principais contribuições deste trabalho:

1. discussão e contextualização dos ambientes de BI como solução para análise e interpretação dos dados no contexto de e-Gov;
2. criação de um ambiente de BI para e-Gov utilizando tecnologias de código aberto, que atende desde o processo de ETL até o provimento de ferramentas para o usuário final;
3. descrição de uma metodologia para guiar o processo de criação do ambiente proposto, que pode facilmente ser estendida para ambientes similares;
4. apresentação de um estudo de caso onde o ambiente de BI proposto foi criado utilizando a metodologia descrita.

O trabalho apresentado atingiu os objetivos propostos. No entanto, torna-se possível o desenvolvimento de trabalhos derivados deste, como:

1. inclusão de ferramentas neste ambiente que gerem análises dos dados de forma automática ou semi-automática, como ferramentas de mineração de dados e análises estatísticas;
2. estudo dos impactos econômicos e sociais da implantação deste ambiente no contexto de e-Gov;
3. aplicação de metodologias e ferramentas para governança das TIC e dados do ambiente de BI;
4. análise de outros atributos de qualidade do ambiente, como disponibilidade, performance das consultas aos dados, acessibilidade, entre outros;

5. criação de um ambiente de BI federado, onde é possível o acesso aos dados de diferentes ambientes de BI pré-existentes de maneira transparente para o usuário.

## 7 Referências Bibliográficas

ALMEIDA, M. S., ISHIKAWA, M., REINSCHMIDT, J., ROEBER, T., **Getting Started with Data Warehouse and Business Intelligence**, IBM Redbooks, 1999, p. 260

AUCOIN, J., **2010 Business Intelligence (BI) challenges and priorities survey**, TechTarget, 2010, Disponível em <<http://searchbusinessanalytics.techtarget.com/news/2240025355/2010-Business-Intelligence-BI-Challenges-and-Priorities-Survey>>

BARZILAI-NAHON, K., SCHOLL, J. H., **Siblings of a different kind: e-Government and e-Commerce**, Proceedings of the 9th IFIP WG 8.5 international conference on Electronic government, 2010, p. 25-37

BÉLANGER, F., CARTER, L., **The impact of the digital divide on e-government use**, Commun. ACM, 2009, v. 52, p. 132-135

BIERE, M. **Business Intelligence for the enterprise**, IBM Press, 2003, p. 240

CARTER, L., BÉLANGER, F., **The Influence of Perceived Characteristics of Innovating on e-Government Adoption**, The Electronic Journal of e-Government, 2004, v. 2, p. 11-20

CHAKRABARTI, S., COX, E., FRANK, E., GTING, R. H., HAN, J., JIANG, X., KAMBER, M., LIGHTSTONE, S. S., NADEAU, T. P., NEAPOLITAN, R. E., PYLE, D., REFAAT, M., SCHNEIDER, M., TEOREY, T. J., WITTEN, I. H., **Data Mining: Know It All**, Morgan Kaufmann Publishers Inc., 2008, p. 477

CHAUDHURI, S., DAYAL, U., **An overview of data warehousing and OLAP technology**, SIGMOD Rec., 1997, v. 26, n. 1, p. 65-74

CHEN, H., **Digital Government: technologies and practices**, Decision Support Systems, 2002, v. 34, p. 223-227

CIBORRA, C., **Interpreting e-government and development: Efficiency, transparency or governance at a distance?**, Information Technology & People, 2005, v.18, n.3, p. 260-279

DASH, S. S., SETHI, I. P. S., **National Do Not Call Registry in India: a step towards restricting unsolicited telemarketing calls**, Proceedings of the 3rd international conference on Theory and practice of electronic governance, 2009, p. 335-340

DAYAL, U., CASTELLANOS, M., SIMITSIS, A., WILKINSON, K., **Data integration flows for Business Intelligence**, EDBT '09: Proceedings of the 12th International Conference on Extending Database Technology, 2009, p. 1-11

DIERKS, T., RESCORLA, E., **The Transport Layer Security (TLS) Protocol Version 1.2**, IETF, 2008, Disponível em <<http://tools.ietf.org/html/rfc5246>>

DUNHAM, M. H., **Data Mining: Introductory and Advanced Topics**, Prentice Hall PTR, 2002, p. 315

EBRAHIM, Z., IRANI, Z., **E-government adoption: architecture and barriers**, Business Process Management Journal, 2005, v. 11, n. 5, p. 589-611

FREIER, A. O., KARLTON, P., KOCHER, P. C., **The SSL Protocol Version 3.0**, Mozilla Foundation, 1996, Disponível em <<http://www.mozilla.org/projects/security/pki/nss/ssl/draft302.txt>>

GALPAYA, H., SAMARAJIVA, R., SOYSA, S., **Taking e-government to the bottom of the pyramid: dial-a-gov?**, Proceedings of the 1st international conference on Theory and practice of electronic governance, 2007, p. 233-241

Gartner Group, **Gartner EXP Worldwide Survey of More than 1,500 CIOs Shows IT Spending to Be Flat in 2009**, 2009, Disponível em <<http://www.gartner.com/it/page.jsp?id=855612>>, Acesso em 16 de set 2010



GORDON, T. F., **Introduction to e-government**, ERCIM News, Special Theme: E-Government, 2002, v.48, p.12-13

HAN, J., KAMBER, M., **Data mining: Concepts and Techniques, Second Edition**, Morgan Kaufmann Publishers Inc., 2006, p. 743

HONG, X., ZAI-WEN, L., HAI-YANG, M., **Study and Realization of Supermarket BI System Based on Data Warehouse and Web Technique**, CSSE '08: Proceedings of the 2008 International Conference on Computer Science and *Software Engineering* , 2008, p. 482-485

IBGE, **Pesquisa Nacional por Amostra de Domicílios**, Disponível em <<http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2009/default.shtm>>, Acesso em 01 de jan 2011

INMON, W. H., **Building the Data Warehouse, 1st Edition**, John Wiley & Sons, Inc., 1992

INMON, W. H., **Building the Data Warehouse, 4th Edition**, John Wiley & Sons, Inc., 2005, p.576

JasperSoft, Disponível em <<http://www.jaspersoft.com/jasperanalysis>>, Acesso em 13 de mar 2010

JPivot, Disponível em <<http://jpivot.sourceforge.net/index.html>>, Acesso em 13 de mar 2010

KALRA, G. and STEINER, D., **Weather Data Warehouse: An Agent-Based Data Warehousing System**, HICSS '05: Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences , 2005, p. 217-226

KEEN, P. G. W., MORTON, M. S. S., **Decision support systems: an organizational perspective**, Addison-Wesley, 1978, p. 264

Keetle, Disponível em <<http://kettle.pentaho.com>>, Acesso em 12 de mar 2010

KIMBALL, R., **The data warehouse toolkit: practical techniques for building dimensional data warehouses**, John Wiley & Sons, Inc., 1996, p. 374

KIMBALL, R., CASERTA, J., **The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data**, John Wiley & Sons, 2004, p. 491

KIMBALL, R., REEVES, L., ROSS, M., THORNWAITE, W., **The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses**, John Wiley & Sons, Inc., 1998, p. 800

KIMBALL, R., ROSS, M., **The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling**, John Wiley & Sons, Inc., 2002, p. 421

LARSON, B. **Delivering Business Intelligence with Microsoft SQL Server 2008**, McGraw-Hill, Inc., 2009, p. 770

LEE, S. M., TAN, X., TRIMI, S., **Current practices of leading e-government countries**. Commun. ACM, 2005, v. 48, n. 10, p. 99-104

LI, S., SHUE, L., LEE, S., **Business Intelligence approach to supporting strategy-making of ISP service management**, Expert Syst. Appl., 2008, v. 35, n. 3, p. 739-754

LUJÁN-MORA, S., TRUJILLO, J., SONG, I., **A UML profile for multidimensional modeling in data warehouses**, Data Knowl. Eng., 2006, v. 59, n. 3, p. 725-769

LUCKEVICH, M., MISNER, S., VITT, E., **Business Intelligence**, Microsoft Press, 2008, p. 220

MARCHIONINI, G., SAMET, H., Brandt, L., **Digital Government**, Commun. ACM, 2003, v. 46, n. 1, p. 24-27

MARTENS, C., **Business Intelligence at age 17**. Disponível em <[http://www.computerworld.com/s/article/266298/BI\\_at\\_age\\_17](http://www.computerworld.com/s/article/266298/BI_at_age_17)> Acesso em 12 de abr de 2009

MENDES, L.S., BOTTOLI, M.L., BREDA, G.D., **Digital cities and open MANs: A new communications paradigm**, LATINCOM '09. IEEE Latin-American Conference on Communications, 2009, p. 1-8

MicroStrategy, Disponível em <<http://www.microstrategy.com.br/>>, Acesso em 13 de março 2010

Mondrian, Disponível em <<http://mondrian.pentaho.org/>>, Acesso em 14 de março 2010

MOSS, L. T., ATRE, S., **Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications**, Addison-Wesley Longman Publishing Co., Inc., 2003, p. 576

MUSSI, C., MURAHOVSKI, D., BETTINI, G., KRATZ, L. G., **Data Warehouse - A Experiência da ANVISA**, IX CBIS - Congresso Brasileiro de Informática em Saúde, 2004, p. 6

MySQL, Disponível em <<http://www.mysql.com/>>, Acesso em 14 de março 2010

Nações Unidas, **United Nations E-Government Survey 2008: From E-Government to Connected Governance**, Nações Unidas, 2008, p. 225, Disponível em <<https://unp.un.org/details.aspx?pid=17515>>

OpenI, Disponível em <<http://openi.org/>>, Acesso em 8 de março 2010

Palo, Disponível em <<http://www.jedox.com/en/home/overview.html>>, Acesso em 14 de março 2010

PANHAN, A. M. IGNATOWICZ, E., MENDES, L. S., **Community portals for architecture-based middleware P2P**, Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government, 2009, p. 5

Prefeitura Municipal de Campinas, <<http://2009.campinas.sp.gov.br/trabalho/>>, Acesso em 13 de março 2010

POWER, D. J., **Decision Support Systems: Concepts and Resources for Managers**, Quorum Books, 2002, p. 225

RESCORLA, E., **HTTP Over TLS**, IETF, 2000, Disponível em <<http://tools.ietf.org/html/rfc2818>>

RIFAIE, M., Blas, E. J., and MUHSEN, A. R. M., MOK, T. T. H., KIANMEHR, K., ALHAJJ, R., RIDLEY, M. J., **Data warehouse architecture for GIS applications**, iiWAS '08: Proceedings of the 10th International Conference on Information Integration and Web-based Applications & Services, 2008, p. 178-185

RIZZI, S., ABELLÓ, A., LECHTENBÖRGER, J., TRUJILLO, J., **Research in data warehouse modeling and design: dead or alive?**, DOLAP '06: Proceedings of the 9th ACM international workshop on Data warehousing and OLAP, 2006, p. 3-10

RUSCHEL, A. J., **Governo eletrônico: *Business Intelligence* para a modernização do Judiciário**, Disponível em <<http://www.buscalegis.ufsc.br/revistas/index.php/buscalegis/article/view/29592>>, 2008, p. 13

SAHAMA, T. R., CROLL, P. R., **A data warehouse architecture for clinical data warehousing**, ACSW '07: Proceedings of the fifth Australasian symposium on ACSW frontiers, 2007, p. 227-232

SANDHU, R. and FERRAILOLO, D. and KUHN, R., **The NIST Model for Role-Based Access Control: Towards a Unified Standard**, NIST, 2010, Disponível em <<http://csrc.nist.gov/rbac/sandhu-ferraiolo-kuhn-00.pdf>>

SERRA, L., **A Essência do *Business Intelligence***, Berkeley, 2002, p. 296

SPRECHER, M. H., **Racing to e-government: Using the Internet for citizen service delivery**, Government Finance Review, 2000, v.16, p. 21-22,

SNELL, J., TIDWELL, D., KULCHENKO, P., **Programming Web services with SOAP**, O'Reilly & Associates, Inc., 2002, p. 264

Talend, Disponível em <<http://www.talend.com>>, Acesso em 8 de março de 2010

TILLI, M., PANHAN, A. M., LIMA, O., MENDES, L. S., **A Web-based Architecture for e-Gov Application Development**, ICE-B: Proceedings of the International Conference on e-Business, 2008, p. 4

VELSEN, L., GEEST, T., HEDDE, M., DERKS, W., **Engineering User Requirements for e-Government Services: A Dutch Case Study**, EGOV '08: Proceedings of the 7th international conference on Electronic Government, 2008, p. 243-254

VERCELLIS, C., *Business Intelligence: Data Mining and Optimization for Decision Making*, John Wiley & Sons, Inc., 2009, p. 448

WAH, T. Y., SIM, O. S., Development of a data warehouse for Lymphoma cancer diagnosis and treatment decision support, WSEAS Trans. Info. Sci. and App., 2009, v. 6, n. 3, p. 530-543

WAGNER, W., ANTONUCCI, Y. L., **An analysis of the imagine PA public sector ERP Project**, Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004, p.

Xmla, Disponível em <<http://www.xmla.org/>>, Acesso em 08 de março 2010

XU, L., ZENG, L., SHI, Z., HE, Q., WANG, M., **Research on Business Intelligence in enterprise computing environment**, ISIC. IEEE International Conference on Systems, Man and Cybernetics, 2007, p. 3270-3275

XU, Z., ZHANG, M., JIANG, X., **Business Intelligence - A Case Study in Life Insurance Industry**, ICEBE '05: Proceedings of the IEEE International Conference on e-Business Engineering, 2005, p. 129-133

YAN, P., GUO, J., **Researching and Designing the Architecture of E-government Based on SOA**, Proceedings of the 2010 International Conference on E-Business and E-Government, 2010, p. 512-515

YU, C. C., **Role-Based and Service-Oriented Security Management in the E-Government Environment**, EGOV '09: Proceedings of the 8th International Conference on Electronic Government, 2009, p. 364-375

ZENG, L., XU, L., SHI, Z., WANG, M., WU, W., **Techniques, Process, and Enterprise Solutions of *Business Intelligence***, SMC '06. IEEE International Conference on Systems, Man and Cybernetics, 2006, p. 4722-4726