

Ferramenta de Construção de Data Warehouse

Percio Alexandre de Oliveira

Prof. Maurício Capobianco Lopes - Orientador

Índice

- Introdução
- Objetivos
- Data Warehouse
 - Estrutura Interna
 - Características
 - Principais elementos: ETC, Metadados e Modelagem Dimensional
- Trabalhos Correlatos
- Requisitos Funcionais e Não Funcionais
- Especificação
 - Diagrama de casos de uso, pacotes, classes e seqüência
- Implementação
- Estudo de caso
- Resultados
- Conclusão
- Extensões



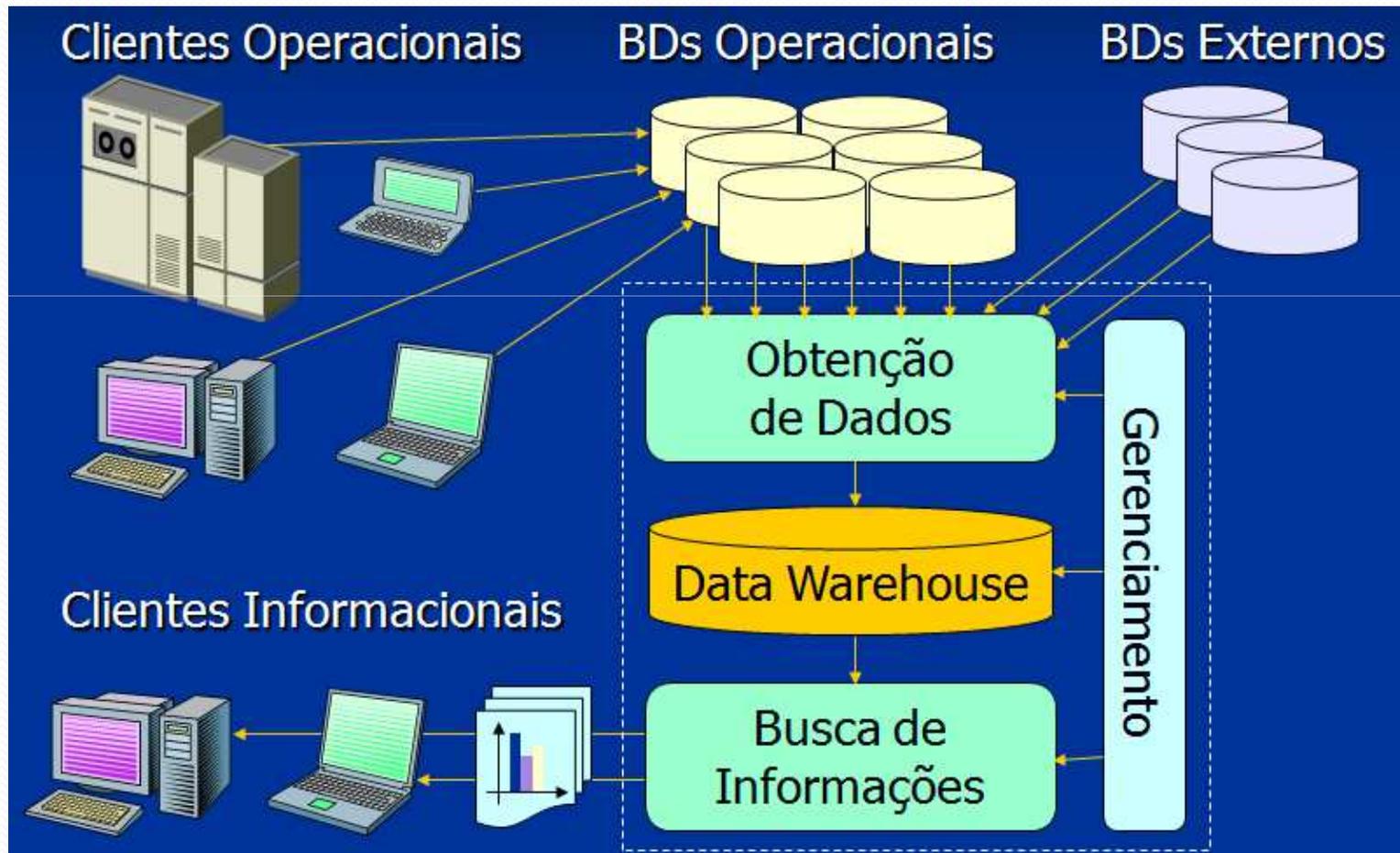
Introdução

- Informação
 - Recurso para tomada de decisões estratégicas
 - Obtida através da análise de dados históricos
 - Dificuldade na obtenção pela grande quantidade de dados a serem analisados
- Análise dos dados
 - Pode aumentar a competitividade da empresa
 - É feita de forma intuitiva pelos gerentes da empresa
 - Ganhar eficiência e lucratividade

Objetivos

- *Transformação de dados históricos provenientes de uma única ou múltiplas fontes em informações sumarizadas e analíticas no auxílio para tomada de decisões.*
- Extrair, Transformar e Carregar dados de múltiplas fontes de dados para um ambiente de data warehouse
- Construir projetos de DW através da técnica de modelagem dimensional Star Schema
- Separar a ferramenta em perfil de usuários e administradores
- Gerar consultas gerenciais com recursos de drill down, ranking, filtros e alertas
- Importar e exportar o metadados

Estrutura Interna



Características

- Dados baseados em assunto
- Ambiente Integrado
- Não-volatilidade
- Variável em relação ao tempo

Extração, Transformação e Carga



Metadados

- Agregam valor aos dados
- Usuários ficam conhecendo a estrutura e o significado dos dados
- Como os dados ?
 - Mudança
 - Sumarizados
 - Acessados
 - Calculados
 - Interpretados
 - Qualificados



Modelagem Dimensional

- Dimensões e atributos
- Tabela Fato e indicadores
- Cubo de decisão
- Modelo Star-Schema



Trabalhos Correlatos

- Oracle
 - Oracle Discoverer
- Microsoft
 - SQL Server 2005 Business Intelligence Development Studio e o SQL Server 2005 Management Studio
 - Excel
- Microsiga
 - SigaDW
- TCCs
 - Laurício Corradi (2002) – Análise na comercialização de planos de previdência
 - Gustavo Jark (2002) – Gerenciamento de clientes/vendas

Requisitos Funcionais

Permitir perfil de usuário e administrador

Permitir administrador criar projetos de DW

Permitir administrador definir conexões com banco SQL e MySql

Permitir administrador definir o processo de ETC

Permitir administrador agendar as cargas de dados

Permitir administrador criar e definir consultas gerenciais

Permitir usuário ou administrador aplicar recursos de alertas, ranking e filtro nas consultas

Permitir administrador de importar ou exportar o metadados

Permitir administrador criar novos usuários

Permitir administrador visualizar *log* de mensagens do DW

Requisitos Não Funcionais

Ambiente WEB

Linguagem JAVA utilizando JSP e Servlets

AJAX para melhor interatividade

Arquitetura dividida em camadas

Independência de plataforma



Especificação

- Diagramas UML
 - Casos de Uso
 - Pacotes
 - Classes
 - Seqüência
- Ferramenta Enterprise Architect

Diagrama de casos de uso

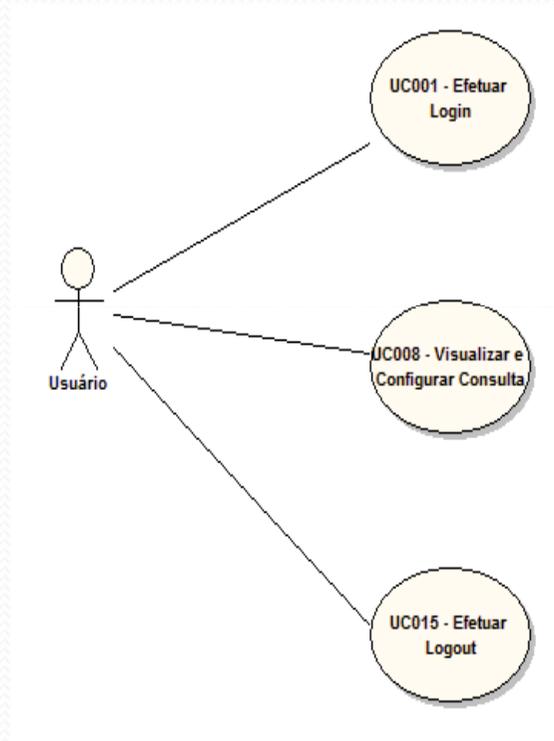


Diagrama de pacotes

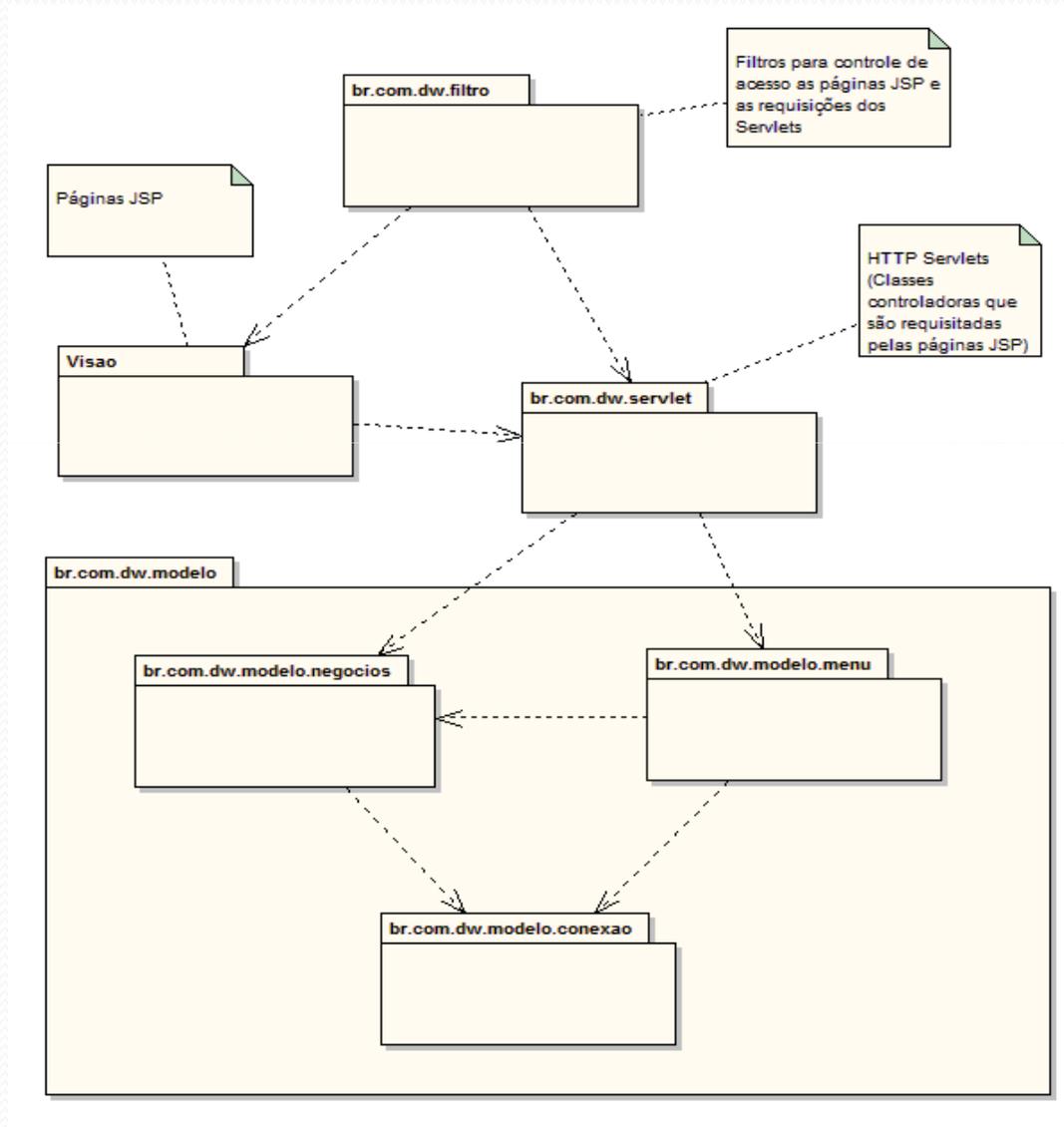


Diagrama de classes - ETC

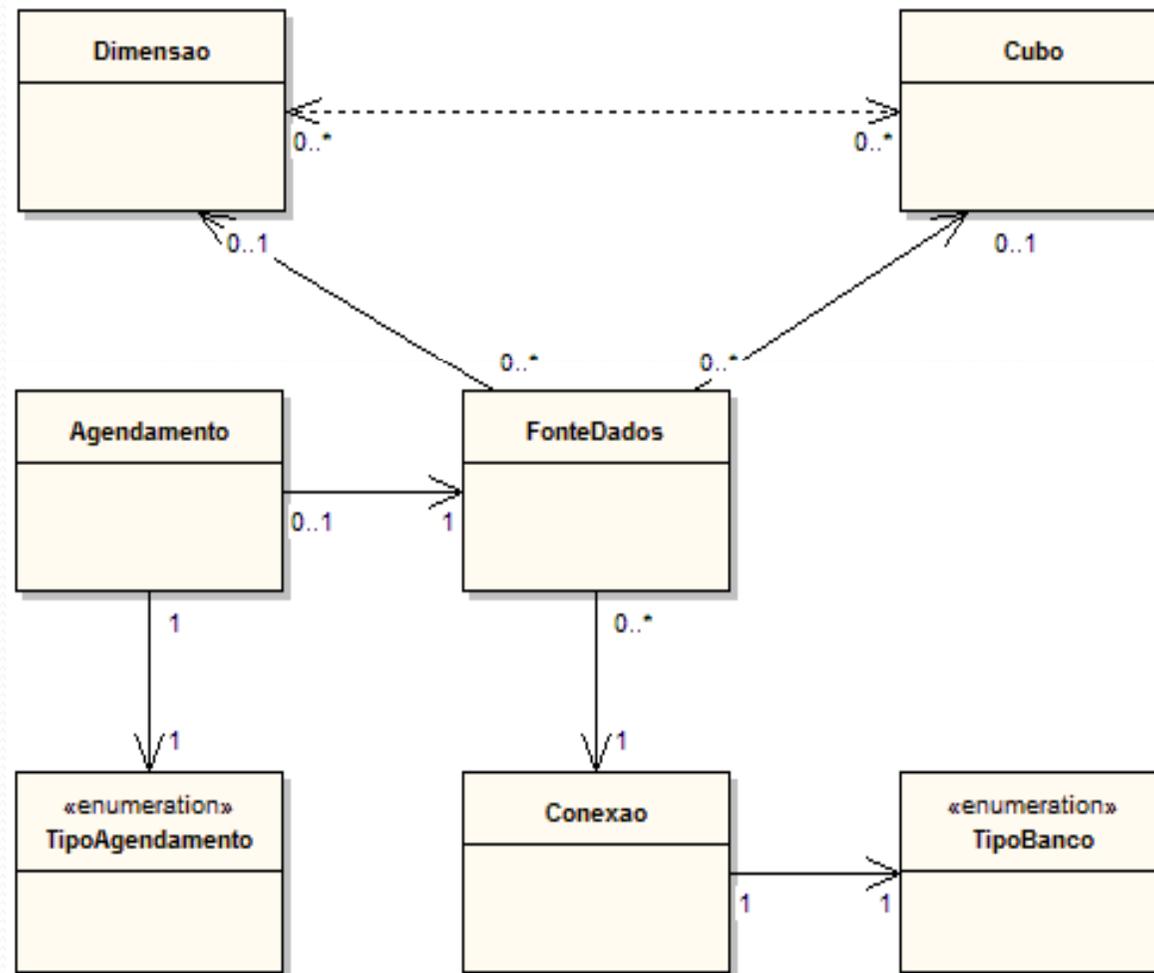


Diagrama de classes - PROJETO

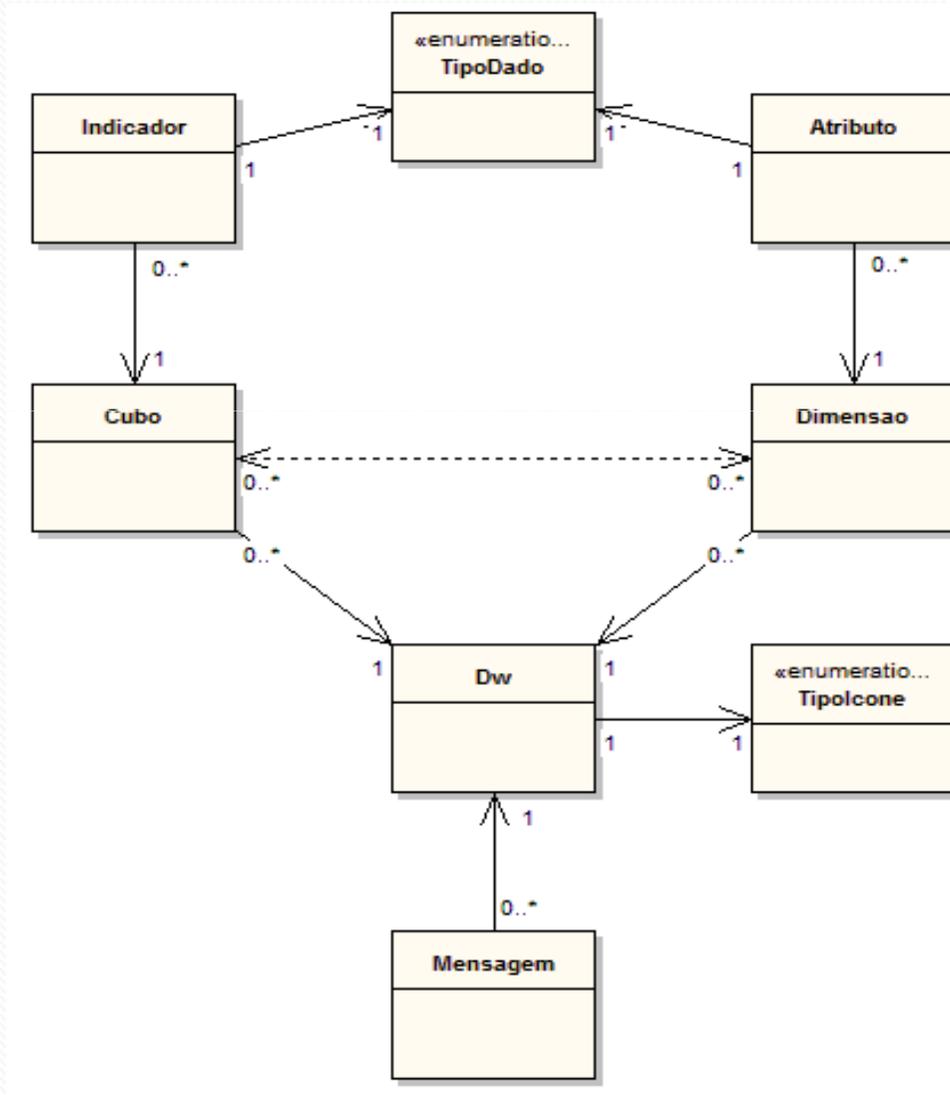


Diagrama de classes - CONSULTA

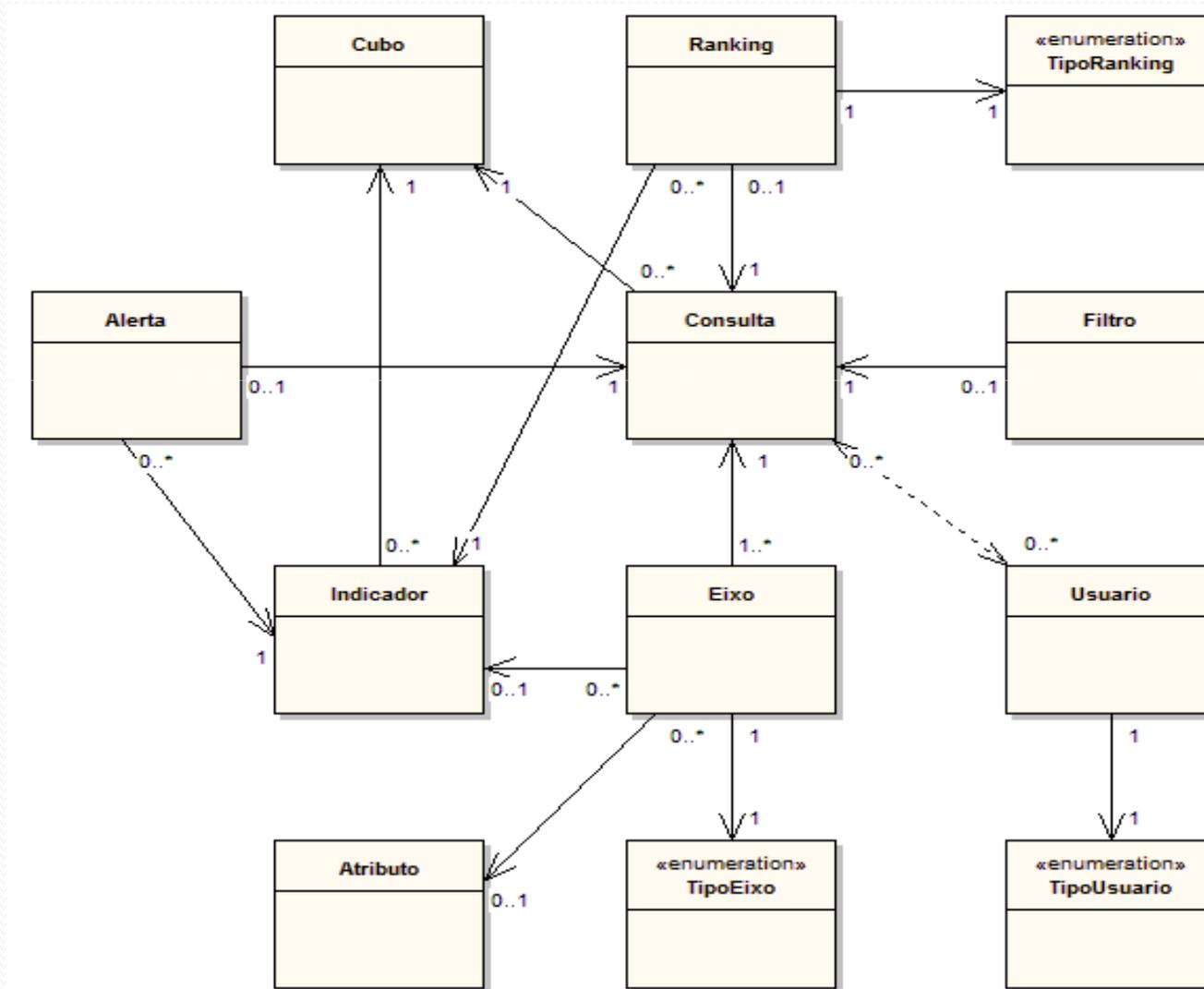


Diagrama de seqüência - ETC

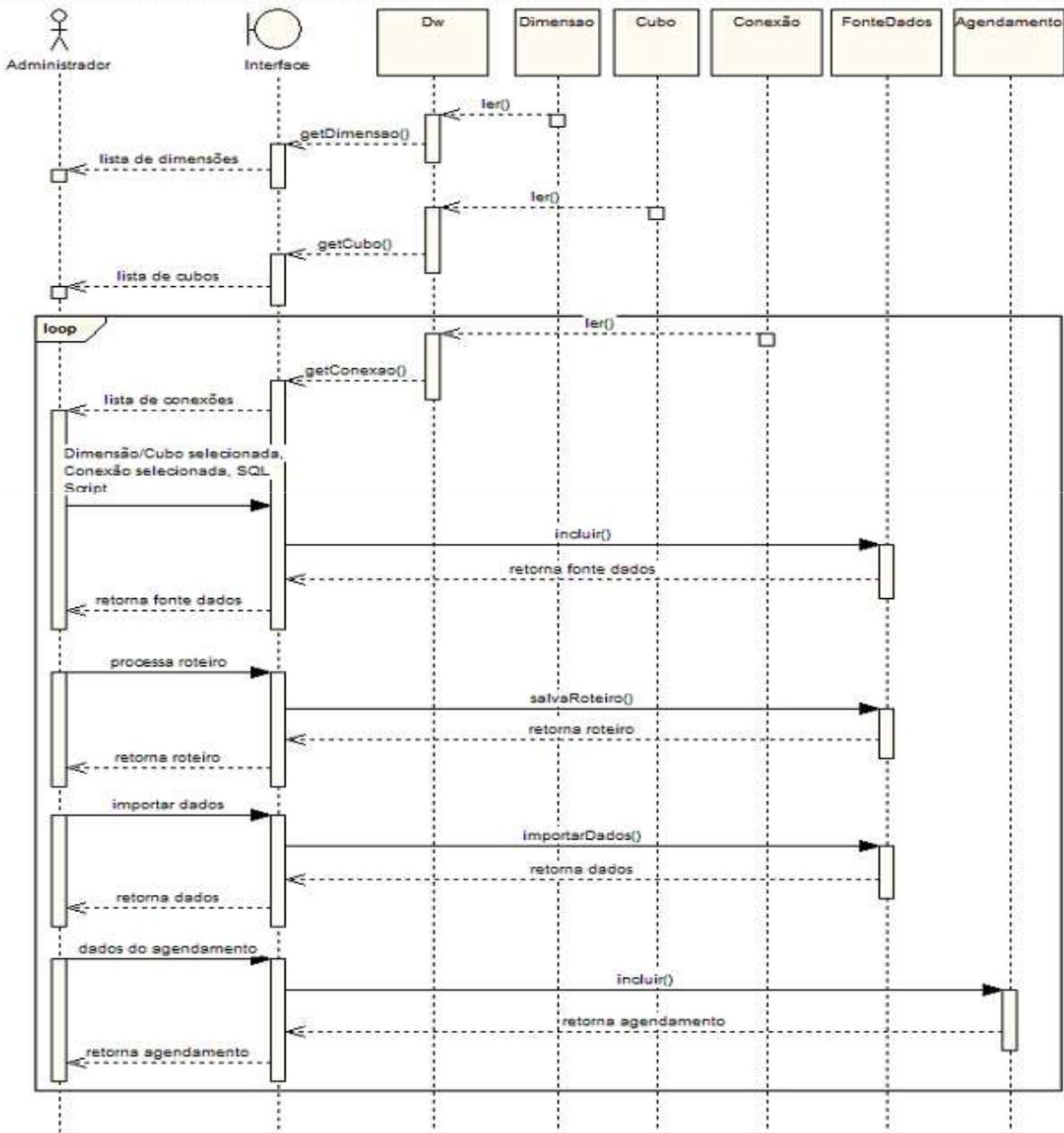


Diagrama de seqüência - PROJETO

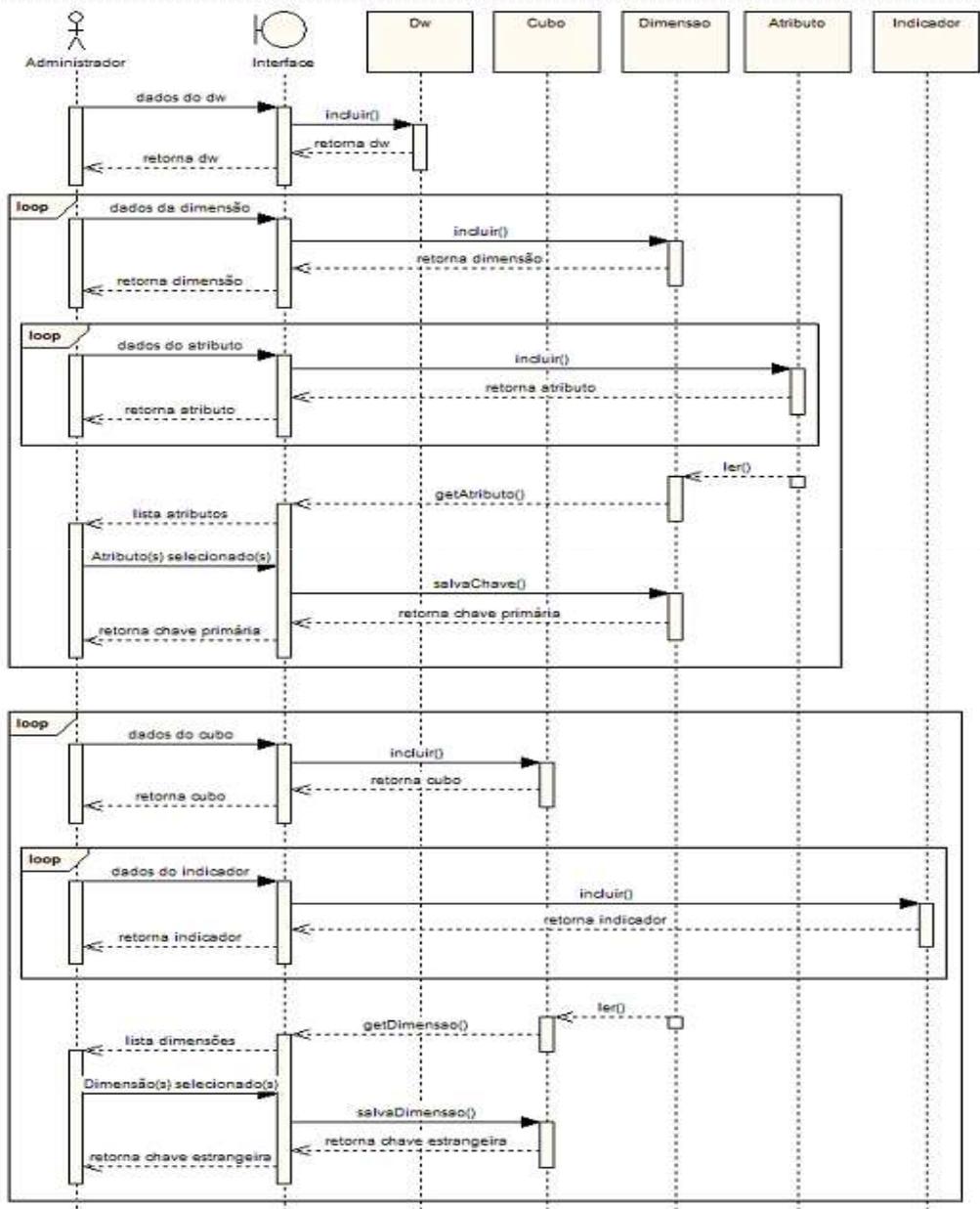
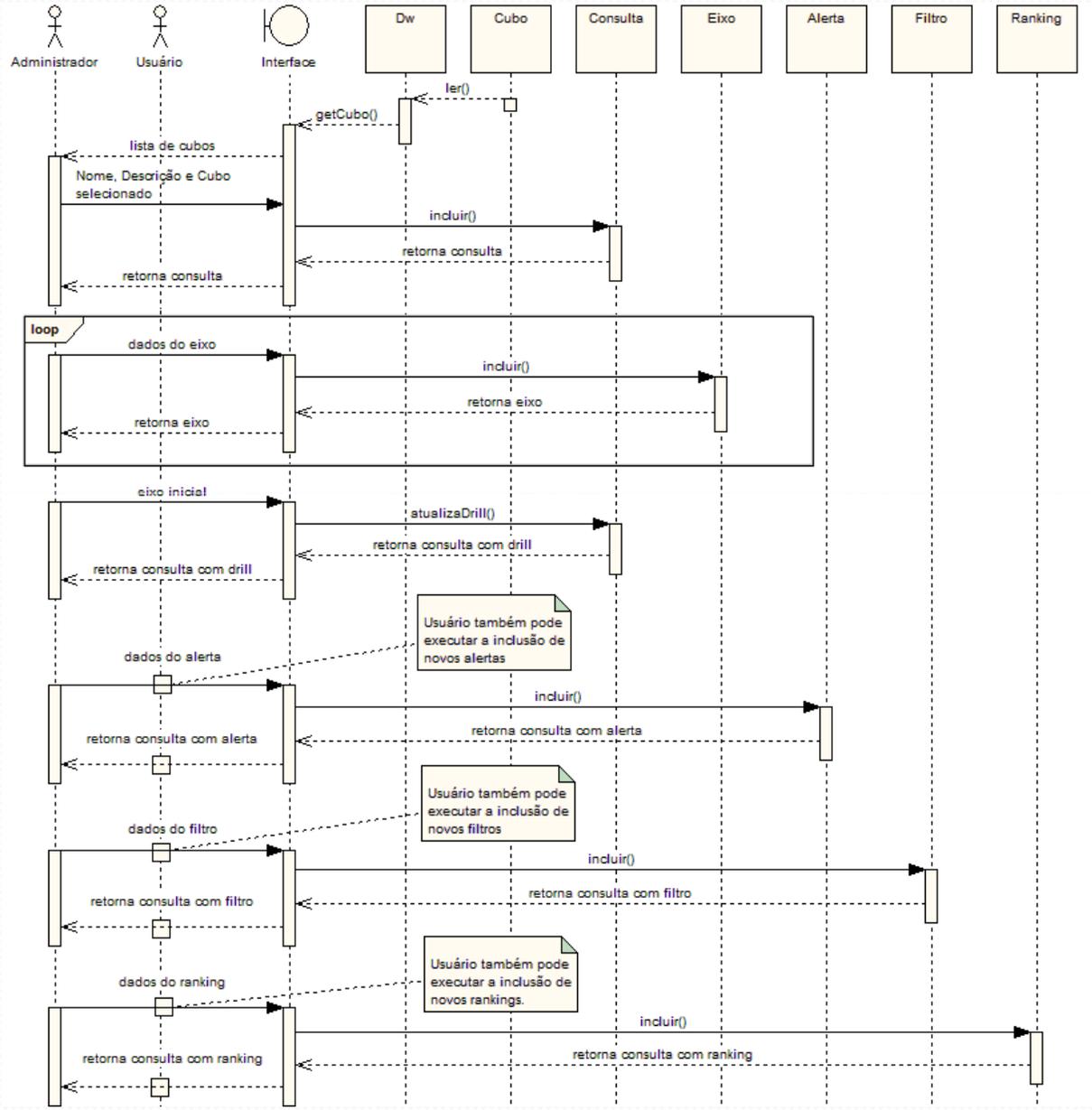


Diagrama de seqüência - CONSULTA



Implementação

- J2SE 1.5
- IDE Eclipse 3.2.1
- Plug-in MyEclipse 5.1.1
- Web container Tomcat 5.5.23
- Biblioteca javascript para montagem de grids da Zapatec
- MySql 5.0 com JDBC
- Javascripts com AJAX

Estudo de caso

- Faturamento da empresa FILLPRINT
- Histórico de vendas
- consultar a soma das vendas em valor total da nota fiscal por data (ano e mês) e clientes (estado)
- consultar a média de vendas em valor total da nota fiscal por representante (nome) e clientes (nome), alertando o usuário onde a média das vendas foram abaixo de duzentos reais na cor vermelha e acima na cor verde
- consultar os vinte produtos mais vendidos em quantidade no ano de 2007

Consulta Ano x Mês

Ferramenta Data Warehouse - Windows Internet Explorer
http://localhost:8080/dw/cadastros/vertabela.jsp?pos=-1&id=1&valor=

ANO VENDA	VALOR TOTAL DA NOTA FISCAL
2004	3573.57
2005	16122.11
2006	134675.83
2007	784410.36

Page 1 of 1 (4 rows)

Done Internet | Protected Mode: Off 100%

Ferramenta Data Warehouse - Windows Internet Explorer
http://localhost:8080/dw/cadastros/vertabela.jsp?id=1&pos=1&valor=2007&vant=2007

ANO VENDA	MES VENDA	VALOR TOTAL DA NOTA FISCAL
2007	1	220440.95
	2	198277.83
	3	179166.66
	4	160096.65
	5	26428.27

Page 1 of 1 (5 rows)

Done Internet | Protected Mode: Off 100%

Consulta Representante x Cliente

Ferramenta Data Warehouse - Windows Internet Explorer
http://localhost:8080/dw/cadastros/vertabela.jsp?pos=-1&id=2&valor=

NOME DO REPRESENTANTE	VALOR TOTAL DA NOTA FISCAL [M]
EDUARDO VIEIRA	787.73
FRANCISCO MARTINS	784.75
JOSE LIMA DA SILVA	413.42

Page 1 of 1 (3 rows)

Done Internet | Protected Mode: Off 100%

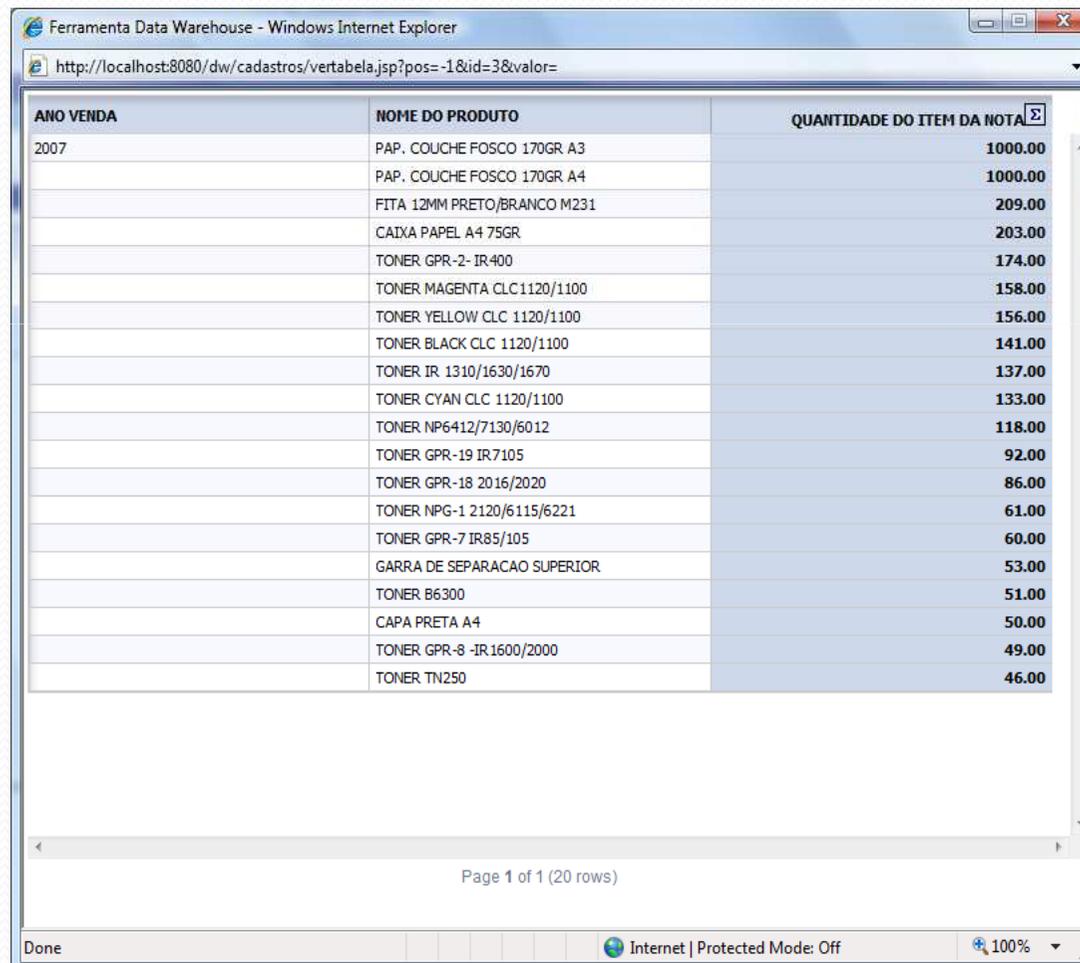
Ferramenta Data Warehouse - Windows Internet Explorer
http://localhost:8080/dw/cadastros/vertabela.jsp?id=2&pos=1&valor=JOSELIMASDASSILVA&vant=JOSELIMASDASSILVA

NOME DO REPRESENTANTE	NOME DO CLIENTE	VALOR TOTAL DA NOTA FISCAL [M]
JOSE LIMA DA SILVA	8A. SUPERINTEND. DE POLICIA RO	58.22
	A4 PRINT SERVICE COM E SERVICO	264.61
	ACIB - ASSOCIACAO EMPRESARIAL	145.24
	ADALBERTO SIDNEI DE MENEZES	129.22
	ADELITA NIEHUES DE AVIS ME	135.73
	ADEMAR DESPACHANTE LTDA	181.55
	ALBANY INTERNATIONAL TECIDOS T	1202.32
	ALBERTO SCHUTZ NETO	470.38
	ALCINO MIRANDA	130.12
	ALFA PAPELARIA LTDA ME	138.93
	ALFA PRINT EDITORA E GRAFICA L	284.28
	ALTENBURG INDUSTRIA TEXTIL LTD	546.81
	AMC TEXTIL LTDA	1068.00
	ANA DA GLORIA DA SILVA MAIA	57.40
	ANA HELENA RIBAS DE ALMEIDA	302.20
	ANA MARIA SCHWERENDT	59.29
	ANA PAULA HOESELMANN	180.60
	ANSELMO VON ZESCHAU	367.68
	APP E.E.B. ERWIN RADTKE	520.00
	APP ESC.MUN.ENS.FUND. ERICH BL	271.46
	APP ESCOLA BASICA MUN. HERMANN	390.44
	APP ESCOLA BASICA MUNIC.PAULIN	124.00
	APP ESCOLA DE ED.BASICA PREF.L	113.48
	AQUARELA DIGITAL LTDA ME	320.98
	ARTELIER DA CASA	258.44

Page 1 2 3 4 5 > >> of 7 (311 rows)

Done Internet | Protected Mode: Off 100%

Consulta Produto



The screenshot shows a web browser window titled "Ferramenta Data Warehouse - Windows Internet Explorer". The address bar contains the URL: <http://localhost:8080/dw/cadastros/vertabela.jsp?pos=-1&id=3&valor=>. The main content area displays a table with three columns: "ANO VENDA", "NOME DO PRODUTO", and "QUANTIDADE DO ITEM DA NOTA". The table lists various products and their quantities for the year 2007. The status bar at the bottom indicates "Done", "Internet | Protected Mode: Off", and "100%".

ANO VENDA	NOME DO PRODUTO	QUANTIDADE DO ITEM DA NOTA
2007	PAP. COUCHE FOSCO 170GR A3	1000.00
	PAP. COUCHE FOSCO 170GR A4	1000.00
	FITA 12MM PRETO/BRANCO M231	209.00
	CAIXA PAPEL A4 75GR	203.00
	TONER GPR-2- IR400	174.00
	TONER MAGENTA CLC 1120/1100	158.00
	TONER YELLOW CLC 1120/1100	156.00
	TONER BLACK CLC 1120/1100	141.00
	TONER IR 1310/1630/1670	137.00
	TONER CYAN CLC 1120/1100	133.00
	TONER NP6412/7130/6012	118.00
	TONER GPR-19 IR7105	92.00
	TONER GPR-18 2016/2020	86.00
	TONER NPG-1 2120/6115/6221	61.00
	TONER GPR-7 IR85/105	60.00
	GARRA DE SEPARACAO SUPERIOR	53.00
	TONER B6300	51.00
	CAPA PRETA A4	50.00
	TONER GPR-8 -IR 1600/2000	49.00
	TONER TN250	46.00

Page 1 of 1 (20 rows)

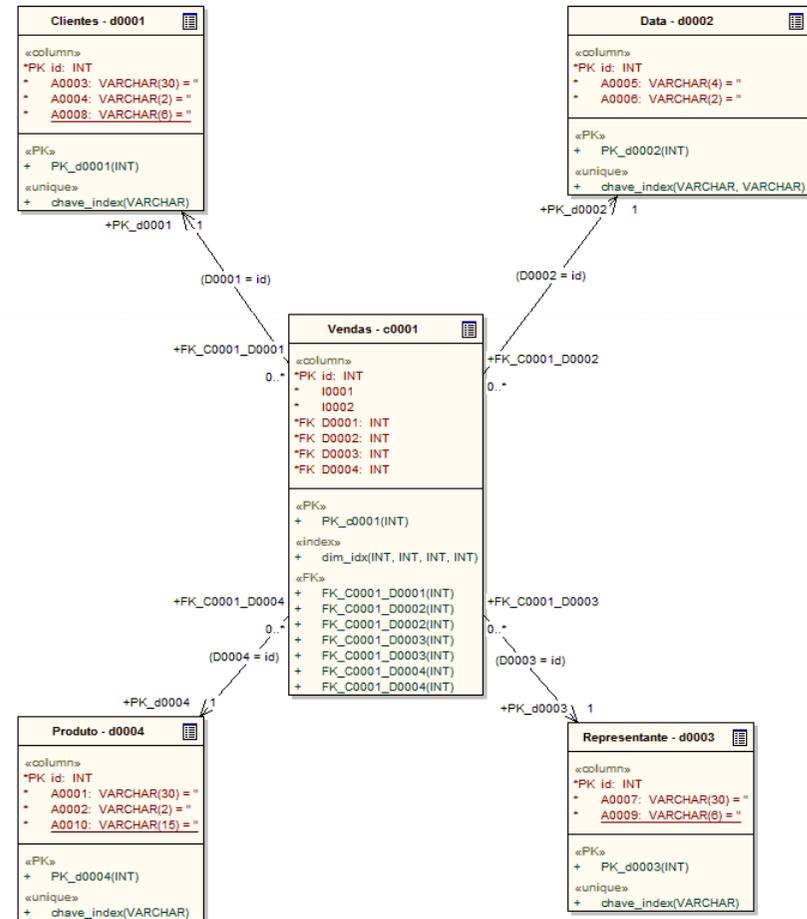
Resultados

Venda ano, mês e estado						
Nível	Tempo processamento (s)		Total de tabelas		Total de registros	
	BD	DW	BD	DW	BD	DW
1	0,0185	0,0064	1	2	3518	2323 x 1
2	0,0192	0,0066	1	2	3518	2323 x 1
3	1,2249	0,0164	2	3	982 x 3518	27 x 145 x 1

Venda representante e cliente						
Nível	Tempo processamento (s)		Total de tabelas		Total de registros	
	BD	DW	BD	DW	BD	DW
1	0,0414	0,0119	2	2	3 x 3518	4 x 232
2	4,3536	0,0145	3	3	3 x 3518 x 982	4 x 232 x 1

Venda produto						
Nível	Tempo processamento (s)		Total de tabelas		Total de registros	
	BD	DW	BD	DW	BD	DW
1	18,1847	0,0195	2	3	3518 x 15969	27 x 145 x 1

	Tempo processamento (min.)	Total de registros
Dimensão Cliente	1m 51s	970
Dimensão Produto	26min 12s	11594
Dimensão Data	5min 23s	3496
Dimensão Representante	< 1s	3
Cubo Venda	6min 45s	4359





Conclusão

- Transformação de dados históricos e operacionais em informação gerencial e estratégica
- Centralização da informação em um banco de dados único com dados integrados, refinados, sumarizados e modelados de acordo com técnicas de modelagem dimensional
- Performance na consulta e obtenção da informação
- Baixo custo
- Acesso a informação através da web, para todo e qualquer usuário que necessite e tenha permissão.



Extensões

- ETC para outros bancos de dados, arquivos e planilhas. Implementação de outros métodos de atualização (incremental, inserção, etc..).
- Consultas: indicadores virtuais, novas funções de agregação (% de participação, média interna, ...), gráficos e rankings (percentual de pareto, curva ABC)
- Integração com ferramentas como Crystal Reports e Excel
- Ferramentas Data Mining para usuário final
- Processamento paralelo