

Data Warehouses: Fundamentos, Ferramentas e Tendências Atuais

Prof. Renato Fileto
Departamento de Informática e Estatística (INE)
Universidade Federal de Santa Catarina (UFSC)

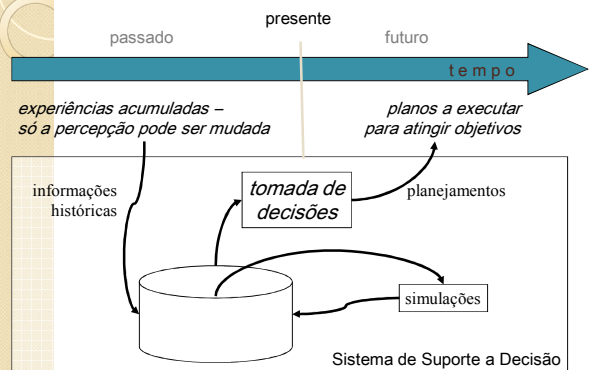
Tópicos

- **Conceitos básicos**
 - Sistemas de Suporte à Decisão
 - Data Warehouses (DWs)
 - OLTP vs. OLAP
 - Modelo Dimensional e operadores OLAP (*drill-down, roll-up, etc.*)
- **Planejando, projetando e implementando DWs**
 - Diretrizes gerais, fases do desenvolvimento e tarefas envolvidas
 - Estudo de Caso: DWs para a agricultura
- **Padrões e ferramentas para a implementação de DWs**
 - Principais componentes e padrões para DWs
 - Ferramentas abertas e proprietárias para DWs
- **Alguns temas de pesquisa atuais em DWs**
 - DWs com extensões espaciais e temporais
 - Semântica em DWs
- **Conclusões e referências para estudos adicionais**

Tópicos

- **Conceitos básicos**
 - Sistemas de Suporte à Decisão
 - Data Warehouses (DWs)
 - OLTP vs. OLAP
 - Modelo Dimensional e operadores OLAP (*drill-down, roll-up, etc.*)
- **Planejando, projetando e implementando DWs**
 - Diretrizes gerais, fases do desenvolvimento e tarefas envolvidas
 - Estudo de Caso: DWs para a agricultura
- **Padrões e ferramentas para a implementação de DWs**
 - Principais componentes e padrões para DWs
 - Ferramentas abertas e proprietárias para DWs
- **Alguns temas de pesquisa atuais em DWs**
 - DWs com extensões espaciais e temporais
 - Semântica em DWs
- **Conclusões e referências para estudos adicionais**

Processos Decisórios



Classes de Sistemas de Informação

- **Sistemas Transacionais**
 - Controlam informações operacionais (por exemplo, vendas, compras, contabilidade, sensoramento e sistemas de tempo real).
 - Operações de manipulação de dados (insert, update, delete), normalmente *on-line* e em nível detalhado.
- **Sistemas de Suporte à Decisão**
 - Extraem informações necessárias para a tomada de decisão, utilizando consultas complexas sobre grandes volumes de dados (por exemplo, determinar a taxa de crescimento do faturamento nos últimos 5 anos).
 - Podem usar sistemas transacionais como fontes de dados.

BDs Transacionais vs. Suporte à Decisão

Característica	BD Transacional	BD Suporte à Decisão
Objetivo	Atividades cotidianas	Análise do negócio
Uso	Operacional	Informativo
Processamento	OLTP	OLAP
Unidade de trabalho	Inclusão, alteração, exclusão	Carga e consulta
Usuários	Operadores (muitos)	Gerência (poucos)
Interação dos usuários	Ações pré-definidas	Pré-definida e <i>ad-hoc</i>
Dados	Operacionais	Analíticos
Volume	Pode ser alto (MB – GB)	Muito alto (GB – TB)
Histórico	60 a 90 dias	vários anos
Granularidade	Detalhada (baixa)	Detalhada e consolidada (alta)
Redundância	Não ocorre (só p/ eficiência)	Pode ocorrer
Estrutura	Estática	Variável
Manutenção	Mínima é o desejável	Constante
Atualização	Contínua (tempo real)	Periódica (<i>snapshots</i> - retratos)
Integridade	Transação	Cada atualização
Acesso a registros	Poucos - por transação	Muitos - para consolidação
Índices	Poucos/simples	Muitos/complexos
Função dos índices	Localizar um registro	Agilizar consultas

Data Warehouse (“Armazém de Dados”)

- Banco de dados voltado para o suporte à tomada de decisão.
- Possivelmente derivado de vários bancos de dados operacionais
- Pode ser usado como base para executar OLAP (On-Line Analytical Processing) e outras tecnologias de análise de informação e extração de conhecimento

Objetivos:

- Satisfazer necessidades de análise de informações
- Monitorar e comparar situações atuais com passadas
- Estimar situações futuras

Definições

• Data Warehouse - DW (W. H. Immon)

- Coleção de dados orientada a assuntos, integrada, com séries temporais e não volátil, voltada para o apoio à tomada de decisão.

• Data Warehousing

- Processo de construção e uso de DWs.

• Business Intelligence (BI)

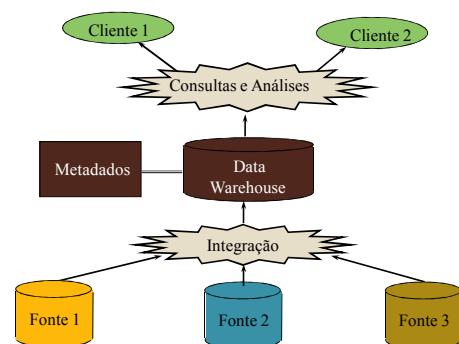
Refere-se a coleta, organização, análise, compartilhamento e monitoramento de informações para suporte a gestão de negócios.

- Inclui Data Warehousing (DW), Data Mining (DM), Customer Relationship Management (CRM).

Características de um DW

- **Orientado a assuntos:** por exemplo, vendas de produtos a diferentes tipos de clientes, atendimentos e diagnósticos de pacientes, rendimento de estudantes
- **Integrado:** diferentes nomenclaturas, formatos e estruturas das fontes de dados precisam ser acomodadas em um único esquema para prover uma visão unificada e consistente da informação
- **Séries temporais:** o histórico dos dados por um período de tempo superior ao usual em BDs transacionais permite analisar tendências e mudanças
- **Não volátil:** os dados de uma data warehouse não são modificados como em sistemas transacionais (exceto para correções), mas somente carregados e acessados para leituras, com atualizações apenas periódicas

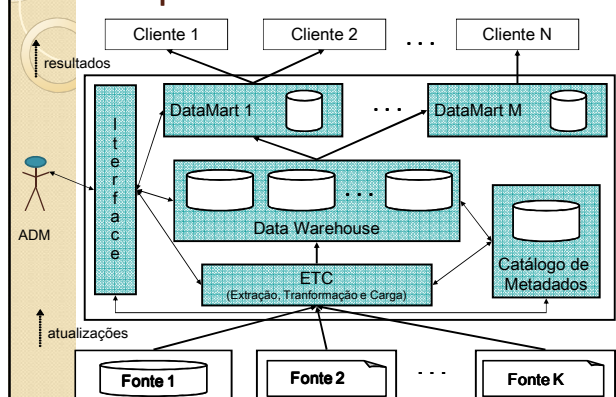
Contexto de Data Warehousing



A Tecnologia de Data Warehouses

- Ferramentas de ETC (Extração, Transformação e Carga) de grande volumes de dados de diversas fontes no DW com recursos para conversão, validação, correção (*data cleansing*) e integração dos dados
- Banco de dados com **modelagem dimensional** voltado para consultas complexas para a obtenção de informação consolidada
- Ferramentas de prospecção e análise de dados baseadas em **OLAP (On-Line Analytical Processing)**
- Ferramentas de administração e gerenciamento do DW e seus **Datamarts (DMs)**

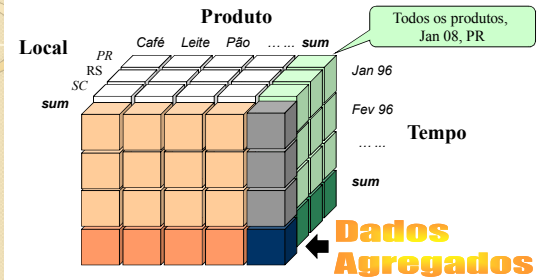
Arquitetura de um DW



O Modelo de dados dimensional

- Modelo específico para processamento analítico de informação (OLAP)
- Medidas organizadas segundo dimensões e suas hierarquias de níveis
 - Exemplos de medidas
 - quantidade vendida
 - valor vendido
 - número de habitantes
 - Exemplos de dimensões
 - **Local** com os níveis país, estado e município
 - **Tempo** com os níveis ano, mês e dia
 - **Produto** com os níveis tipo e nome

O Modelo de dados dimensional



- Células ordinárias (brancas) têm dados no nível mínimo de granularidade para todas as dimensões
- Faces coloridas com dados agregados (count, sum, max, etc.) nas respectivas dimensões

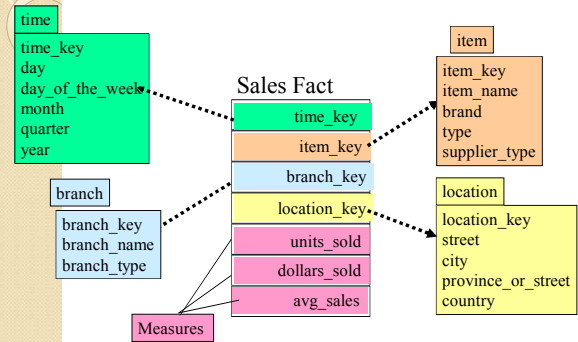
O Esquema de um Data Warehouse

- **Tabela(s) fato** – *Dados quantitativos* – registros de medidas, com dados integrados de várias fontes (muitos registros)
- **Dimensões** – *Dados qualitativos* - organizando conceitos e respectivas instâncias para a seleção e agregação dos dados quantitativos, rotulando esses dados e os resultados (poucos registros)

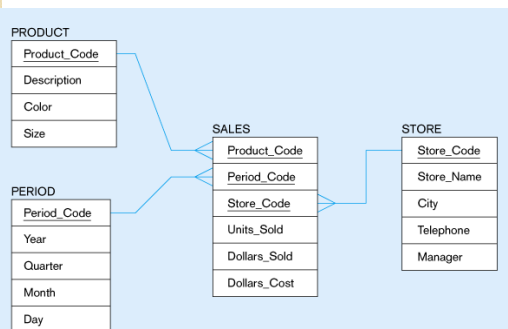
Modelagem de dados em data warehouses:

- *Star* (modelo em formato estrela)
- *Snowflake* (formato de floco de neve)
- *Hypercube* (modelagem em hiper-cubo)

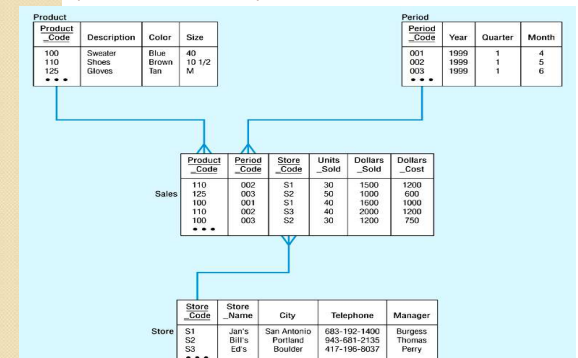
Um esquema em formato estrela



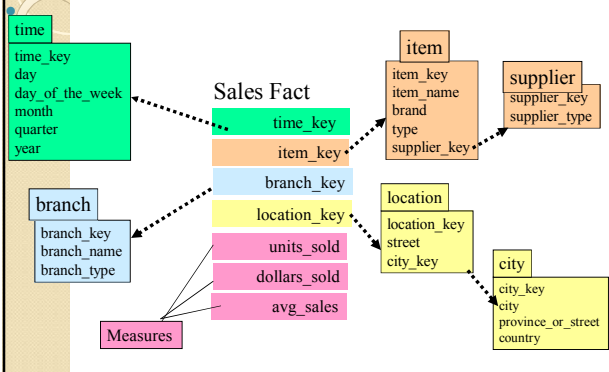
Cardinalidades do esquema estrela



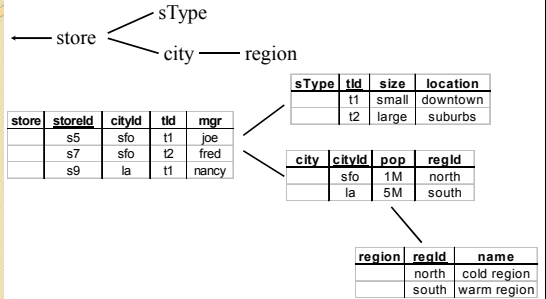
Dados em um esquema estrela (não normalizado)



Um esquema em formato floco de neve



Hierarquias de dimensões (podem ser normalizadas)



Uma constelação de Tabelas Fato

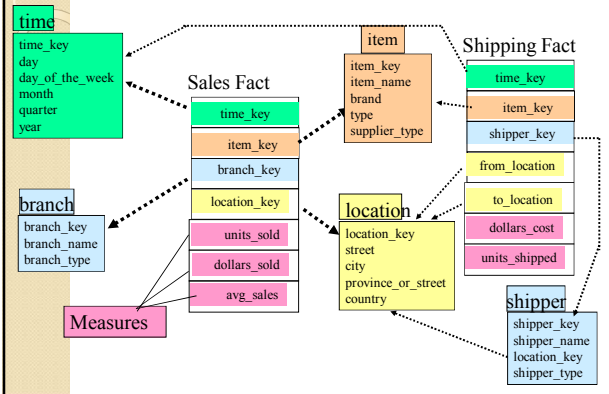


Tabela fato para cubo 2D

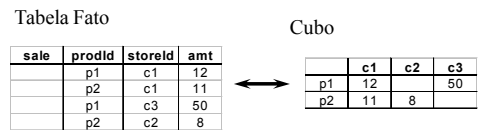
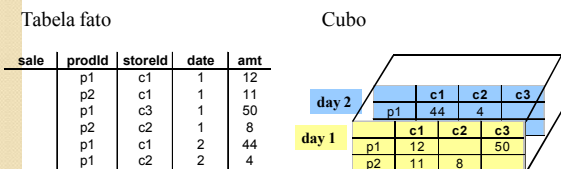


Tabela fato para cubo 3D



Agregação de dados

• Quantidade vendida no dia 1
 SELECT sum(amt)
 FROM SALE
 WHERE date = 1

sale	proddid	storeid	date	amt
	p1	c1	1	12
	p2	c1	1	11
	p1	c3	1	50
	p2	c2	1	8
	p1	c1	2	44
	p1	c2	2	4

➔ 81

Agregação de dados (II)

- **Quantidade vendida por dia**
 SELECT date, sum(amt)
 FROM SALE
 GROUP BY date

sale	prodId	storeId	date	amt
	p1	c1	1	12
	p2	c1	1	11
	p1	c3	1	50
	p2	c2	1	8
	p1	c1	2	44
	p1	c2	2	4



ans	date	sum
	1	81
	2	48

25

Agregação de dados (III)

- **Quantidades vendidas por produto e dia**
 SELECT prodId, date, sum(amt)
 FROM SALE
 GROUP BY date, prodId

sale	prodId	storeId	date	amt
	p1	c1	1	12
	p2	c1	1	11
	p1	c3	1	50
	p2	c2	1	8
	p1	c1	2	44
	p1	c2	2	4



sale	prodId	date	amt
	p1	1	62
	p2	1	19
	p1	2	48

rollup →

← drill-down

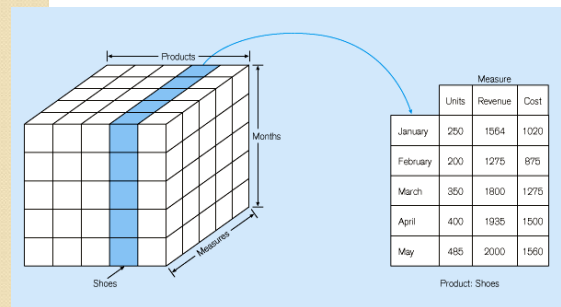
26

Operadores OLAP

- **Slice:** Projeta valores específicos de uma dimensão (extrai uma fatia do hiper-cubo)
- **Dice:** Slices consecutivos (extrai hiper-cubo menor)
- **Roll-up (drill-up):** sumariza dados, subindo na hierarquia de uma dimensão
- **Drill-down (roll-down):** reverso de roll-up, isto é, detalha os dados, descendo na hierarquia de uma dimensão
- **Pivot:** muda posição ou orientação da dimensões na projeção bidimensional de dados do hiper-cubo

27

Slice



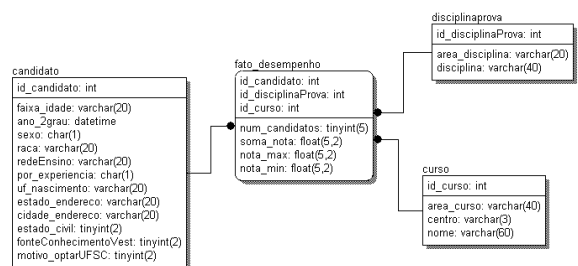
28

Drill-down

Brand	Package size	Sales
Softowel	2-pack	\$75
Softowel	3-pack	\$100
Softowel	6-pack	\$50

Brand	Package size	Color	Sales
Softowel	2-pack	White	\$30
Softowel	2-pack	Yellow	\$25
Softowel	2-pack	Pink	\$20
Softowel	3-pack	White	\$50
Softowel	3-pack	Green	\$25
Softowel	3-pack	Yellow	\$25
Softowel	6-pack	White	\$30
Softowel	6-pack	Yellow	\$20

Exemplo: DW vestibular UFSC (Felipe Shigunov, UFSC, 2007)



Visualização no Openl (Felipe Shigunov, UFSC, 2007)

Curso	Candidato	disciplinaprova	Measures		
			Num Candidatos	maxima	Media notas
-All Cursos	+All Candidatos	+All disciplinas	26.538	10,0	3,8
+CCA	+All Candidatos	+All disciplinas	732	9,8	3,6
+CCB	+All Candidatos	+All disciplinas	850	10,0	3,8
+CCE	+All Candidatos	+All disciplinas	2.852	10,0	3,7
+CCJ	+All Candidatos	+All disciplinas	2.051	10,0	4,0
+CCS	+All Candidatos	+All disciplinas	5.516	10,0	4,0
+CDS	+All Candidatos	+All disciplinas	801	9,6	3,3
+CED	+All Candidatos	+All disciplinas	590	9,5	3,2
+CFH	+All Candidatos	+All disciplinas	2.303	10,0	3,6
+CFM	+All Candidatos	+All disciplinas	796	10,0	3,6
+CSE	+All Candidatos	+All disciplinas	3.195	9,8	3,5
		-All disciplinas	6.852	10,0	3,9
		+Ciencias Biologicas	1.142	9,8	4,8
		+Ciencias Exatas	1.142	10,0	3,7
		+FISICA	571	9,8	3,8
		+MATEMÁTICA	571	10,0	3,6
+CTC	+All Candidatos	+Ciencias Sociais	4.568	9,8	3,8

Drill Down

- É usado para solicitar uma visão mais detalhada de um conjunto de dados. Pode-se dizer que o usuário "mergulha" nos dados.

Candidato	Measures	
	maxima	num
+All Candidatos	10,00	318,468

Candidato	Measures	
	maxima	num
+All Candidatos	10,00	318,468
+Amapa	6,75	12
+Bahia	9,49	324
+Goias	9,80	1,248
+Mato Grosso	10,00	1,548
+Mato Grosso do Sul	9,80	2,256
+Minas Gerais	9,83	1,704
+Parana	10,00	17,208
+Rio de Janeiro	9,50	552
+Rio Grande do Sul	9,67	10,752
+Santa Catarina	10,00	250,128
+Sao Paulo	10,00	29,484

Drill-Down

Roll Up

- Consiste na operação inversa ao Drill-Down, ou seja, apresenta os dados cada vez mais agrupados ou sumarizados.

Candidato	Measures	
	maxima	num
+All Candidatos	10,00	318,468

Candidato	Measures	
	maxima	num
+All Candidatos	10,00	318,468
+Amapa	6,75	12
+Bahia	9,49	324
+Goias	9,80	1,248
+Mato Grosso	10,00	1,548
+Mato Grosso do Sul	9,80	2,256
+Minas Gerais	9,83	1,704
+Parana	10,00	17,208
+Rio de Janeiro	9,50	552
+Rio Grande do Sul	9,67	10,752
+Santa Catarina	10,00	250,128
+Sao Paulo	10,00	29,484

Roll Up

Pivoting

- Serve para adicionar ou rearranjar as dimensões das tabelas

Curso	Candidato	Measures	
		num	num
+All cursos	+Alogos	84	
	+Amapa		12
	+Bahia		324
	+Goias		1,248
	+Maranhao		12
	+Parana		17,208
	+Sao Paulo		29,484

Candidato	Curso	Measures	
		num	num
+All cursos	+All cursos	84	
+Amapa	+All cursos		12
+Bahia	+All cursos		324
+Goias	+All cursos		1,248
+Maranhao	+All cursos		12
+Parana	+All cursos		17,208
+Sao Paulo	+All cursos		29,484

Pivot

Slice and Dice

- Para fixar uma informação de dimensão ou reduzir as dimensões de apresentação dos dados

Curso	Candidato	Measures	
		num	num
+All cursos	-All Candidatos	318,468	
	+Bahia		324
	+Mato Grosso		1,548
	+Minas Gerais		1,704
	+Parana		17,208
	+Rio de Janeiro		552
	+Santa Catarina		250,128
	+Sao Paulo		29,484

Slice and Dice

Tópicos

- **Conceitos básicos**
 - Sistemas de Suporte à Decisão
 - Data Warehouses (DWs)
 - OLTP vs. OLAP
 - Modelo Dimensional e operadores OLAP (*drill-down*, *roll-up*, etc.)
- **Planejando, projetando e implementando DWs**
 - Diretrizes gerais, fases do desenvolvimento e tarefas envolvidas
 - Estudo de Caso: DWs para a agricultura
- **Padrões e ferramentas para a implementação de DWs**
 - Principais componentes e padrões para DWs
 - Ferramentas abertas e proprietárias para DWs
- **Alguns temas de pesquisa atuais em DWs**
 - DWs com extensões espaciais e temporais
 - Semântica em DWs
- **Conclusões e referências para estudos adicionais**

Projeto e Implementação de DWs

- Data warehouses podem requerer muito investimento (tempo, dinheiro, trabalho), dificultando sua adoção em instituições de pequeno e médio porte.
- A modelagem é crítica para o sucesso de uma data warehouse e merece atenção.
- Empreendimentos que não considerem as diferenças entre modelagem de bancos de dados convencionais e data warehouses, incluindo questões técnicas e administrativas, podem facilmente fracassar.



Necessidade de critérios para planejamento, projeto e implementação de data warehouses

Critérios para projeto de DW

- **Determine um escopo pequeno**
- **Escolha um departamento**
- **Defina com clareza os objetivos**
- **Utilize os recursos tecnológicos disponíveis**
- **Não proponha um projeto corporativo**
- **Conceba um projeto escalável**

Formas de Desenvolvimento de DWs

- **Top-down:** Projeto e implementação do DW completo definindo o esquema integrado, fontes de dados e Datamarts
- **Bottom-up:** Projeto e implementação de pequenas DWs ou DMs que vão se integrando aos poucos
- **Combinada:** Mistura desenvolvimento de DWs com várias fontes de dados e com diversos DMs com integração incremental

Fases do desenvolvimento de DWs

1. Planejamento
2. Levantamento das necessidades e fontes de dados
3. Integração de dados
4. Modelagem dimensional
5. Projeto físico do banco de dados
6. Projeto das transformações de dados (ETC)
7. Desenvolvimento de aplicações
8. Validação e teste
9. Treinamento
10. Implantação

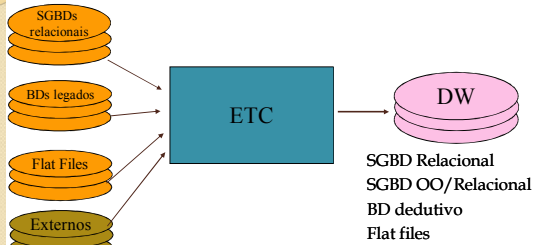
Modelagem Dimensional

- É crítica para o sucesso de uma DW
- É diferente da modelagem de dados convencional
 - A forma como o usuário visualiza e manipula os dados (organização em hiper cubo) deve prevalecer
 - A implementação pode ser em SGBDs específicos para DW ou relacionais/convencionais (verificar a forma como são realizadas junções e outras operações)
 - Diagramas em estrela e floco de neve são utilizados para a implementação de bancos de dados em hiper cubos sobre o modelo relacional
 - Normalização pode ser dispensada, especialmente nas dimensões, por questões de eficiência
 - Abordagem *top-down*

Passos da Modelagem Dimensional

1. Definir a área de negócios (prioridades, mercado, custos e benefícios)
2. Definir os processos dentro da área de negócios
3. Determinar a granularidade desejada (e viável)
4. Definir a(s) tabela(s) fato
5. Descrever as dimensões
6. Definir as métricas para as medidas
 - **Aditivas:** faz sentido adicionar (e.g., valor)
 - **Semi-Aditivas:** faz sentido somar em certas dimensões (e.g., qtde. vendida no tempo/espaco, qtde. de chuva só no tempo)
 - **Não Aditivas:** não faz sentido somar (e.g., valor-venda/custo)
7. Escolher um *DataMart* (definido por uma tabela fato e as dimensões associadas, para iniciar o desenvolvimento)

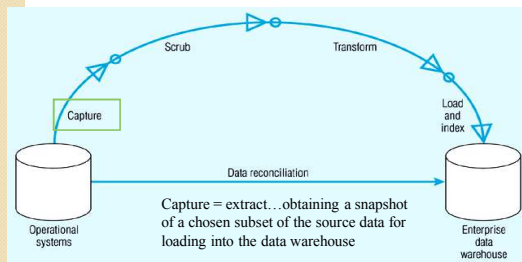
Extração, Transformação e Carga de Dados (ETC – ETL em inglês)



Tarefas de ETC

- **Filtragem de dados:** para eliminar erros e elementos indesejados
- **Integração de dados:** correlaciona dados de fontes heterogêneas com os da data warehouse
- **Conversão de dados:** procedimentos para transformação entre formatos e unidades
- **Condensação de dados:** para reduzir o volume e/ou agilizar o processamento
- **Derivação de dados:** define fórmulas para produzir novos valores a partir dos existentes

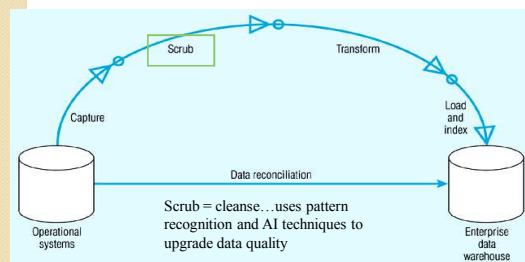
ETL Steps



Static extract = capturing a snapshot of the source data at a point in time

Incremental extract = capturing changes that have occurred since the last static extract

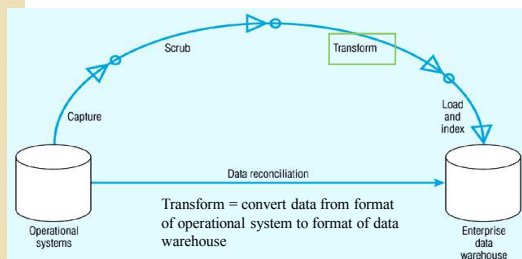
ETL Steps (II)



Fixing errors: misspellings, erroneous dates, incorrect field usage, mismatched addresses, missing data, duplicate data, inconsistencies

Also: decoding, reformatting, time stamping, conversion, key generation, merging, error detection/logging, locating missing data

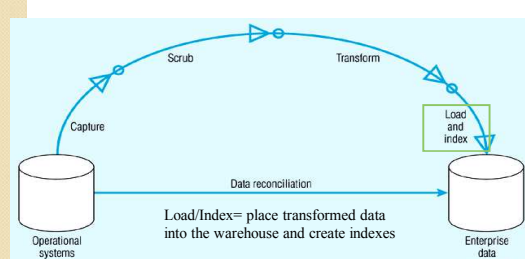
ETL Steps (III)



Record-level:
Selection – data partitioning
Joining – data combining
Aggregation – data summarization

Field-level:
single-field – from one field to one field
multi-field – from many fields to one, or one field to many

ETL Steps (IV)



Refresh mode: bulk rewriting of target data at periodic intervals

Update mode: only changes in source data are written to data warehouse

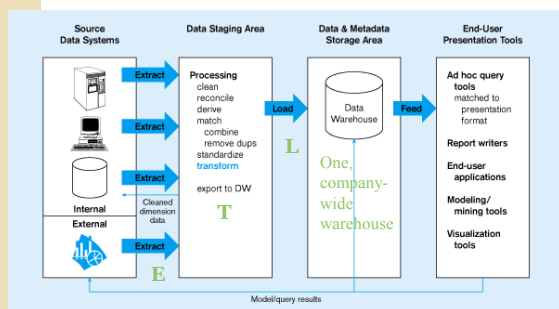
Qualidade de Dados

- **Precisão:** grau de correção (e.g., casas decimais)
- **Consistência:** grau de coerência entre os dados
- **Integridade:** não falta pedaços do banco de dados e não há dados corrompidos
- **Abrangência:** escopo do banco de dados como um todo (e.g., intervalo de tempo, espaço geográfico coberto, gama de produtos)
- **Granularidade:** escopo mínimo de cada valor de dado (e.g., valores totais, mínimos e/ou máximos para cada hora ou cada dia)

Localização dos dados de um DW

- **Repositório centralizado (global):** solução muito utilizada, mas com o inconveniente de requerer investimento em um servidor com alta capacidade de processamento e armazenamento.
- **Repositórios Independentes:** dados armazenados em diferentes locais ou DMs de acordo com áreas de interesse (e.g., financeiro, marketing).
- **Níveis de Detalhes:** dados altamente consolidados/resumidos em um servidor e dados detalhados em outro.
- **Datamarts lógicos:** DMs são visões lógicas dos dados integrados no DW.

DW Centralizado

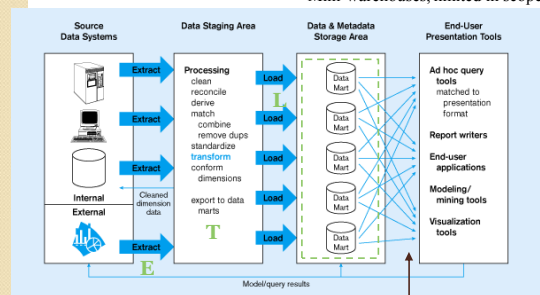


51

DW Descentralizado

Data marts:

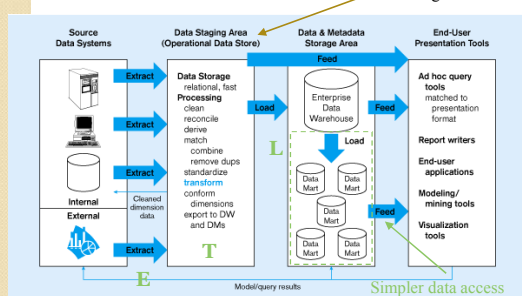
Mini-warehouses, limited in scope



Separate ETL for each independent data mart

Data access complexity due to multiple data marts

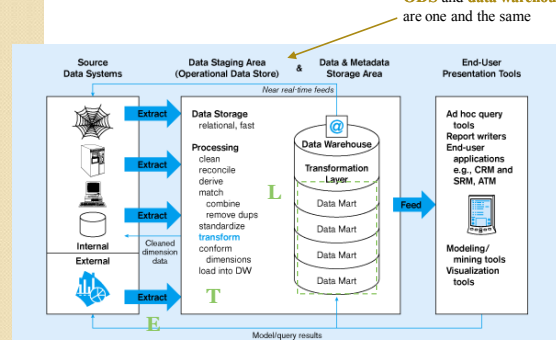
ODS provides option for obtaining current data



Single ETL for enterprise data warehouse (EDW)

Dependent data marts loaded from EDW

ODS and data warehouse are one and the same



Near real-time ETL for @active Data Warehouse

Data marts are NOT separate databases, but logical views of the data warehouse
→ Easier to create new data marts

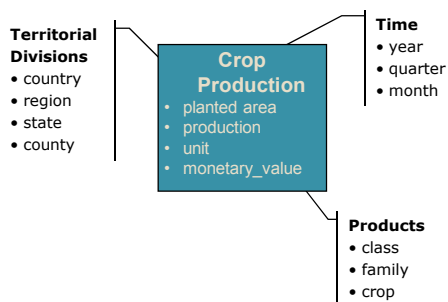
A equipe de desenvolvimento de um DW

Função	Responsabilidades
Gerente da data warehouse	Definição das estratégias, planejamento e gerenciamento
Arquiteto de dados	Análise dos requisitos de dados, modelagem de dados e definição de visões
Administrador de metadados	Definição dos padrões de metadados e gerenciamento do catálogo de metadados
Administrador do BD	Criação das estruturas físicas e monitoramento da carga de dados e desempenho das consultas
Usuário de nível gerencial	Descrição dos dados necessários, especificação das regras do negócio e teste dos resultados
Analista de suporte à decisão	Desenvolvimento de aplicações de suporte à decisão
Especialista em aplicações OLTP	Ajuda na localização dos dados e uso dos recursos de consulta (<i>dump</i>) nos sistemas transacionais
Analista e programador de conversões	Indicação e catalogação das fontes de dados e desenvolvimento de programas para carga de dados
Especialista em suporte técnico	Atividades técnicas como instalação e configuração de hardware e software
Instrutor	Treinamento dos usuários para utilização da data warehouse

Estudo de Caso: DWs para a Agricultura

- Projetos realizados junto à Embrapa, Epagri, Unicamp e Georgia Institute of Technology
 - Análise da produção agrícola
 - Agroclimatologia
 - Zoneamento Agrícola
- Ilustram aplicações de conceitos e técnicas fundamentais
 - Conceitos básicos
 - Modelagem dimensional (e.g., hierarquias complexas)
 - Operadores de agregação de dados
 - Integração de dados
- Demonstram a necessidade de extensões
 - Espaciais
 - Temporais

Agricultural Production Estimation



DW Produção de Frutas no Brazil (Carlos Meira, Embrapa, 2003)

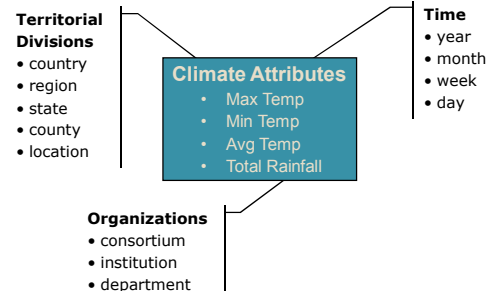
Product	Local	Planted Area (ha)		Production		Unity
		2001	2002*	2001	2002*	
Orange	Brazil	825.228	828.437	16.983.436	18.931.919	tons
	Center	9.289	9.921	131.289	145.866	
	North	18.280	16.724	252.317	233.539	
	North-East	109.584	111.233	1.530.322	1.731.698	
	South	52.003	49.210	795.326	740.559	
	South-East	636.072	641.349	14.250.578	16.080.257	
	Espirito Santo	2.735	2.752	29.343	29.907	
	Minas Gerais	43.895	43.418	575.590	599.999	
	Rio de Janeiro	7.955	7.121	115.753	104.501	
São Paulo	581.487	588.058	13.529.892	15.345.850		
Banana	Brazil	510.313	523.757	6.177.293	6.455.067	tons
Coconut	Brazil	275.551	273.306	1.420.547	1.811.773	10 ³ fruits
Pineapple	Brazil	63.282	64.150	1.468.897	1.450.033	10 ³ fruits
Papaya	Brazil	30.733	31.080	722.986	857.824	tons

* Estimation by December 2002 (actual production data not available yet)

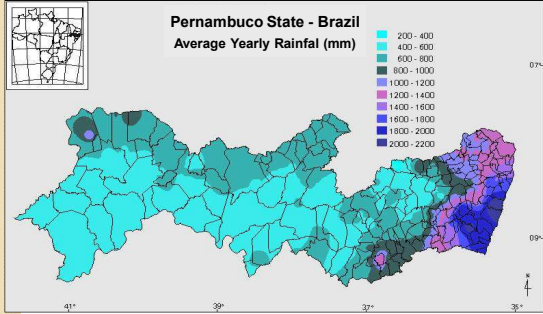
DW Produção Agrícola (Renato Deggau, Epagri/UFSC, 2009)

Produtos	Tempo	Local	Soma Área Plantada
+ All Produtos	+ All Tempo	+ All Local	63.682
Cebola	2007	+ Norte	0
		+ Nordeste	15.934
		+ Centro Oeste	1.348
		+ Sudeste	7.788
		+ MG	1.534
		+ SP	6.125
		+ RJ	0
		+ ES	129
		+ Sul	38.612
		+ PR	6.653
		+ SC	20.795
		+ Oeste Catarinense	2.135
		+ Norte Catarinense	412
		+ Serrana	2.082
		+ Vale do Itajaí	9.689
		+ Grande Florianopoli	6.381
		+ Sul Catarinense	96
		+ RS	11.164

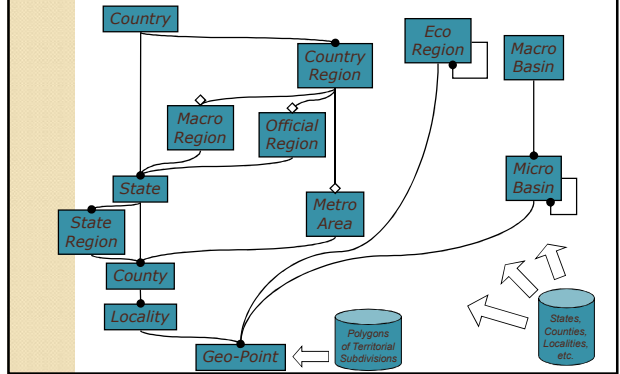
Climate Data



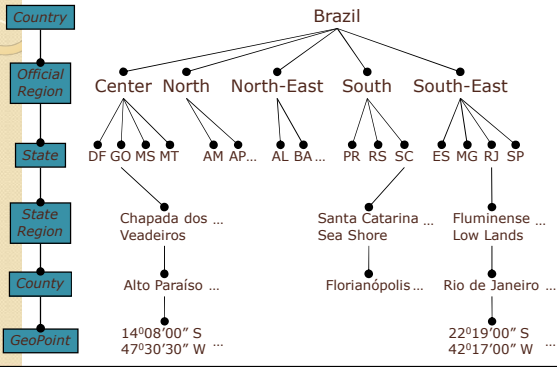
Average Rainfall in Pernambuco State (Cepagri/Unicamp, 2002)



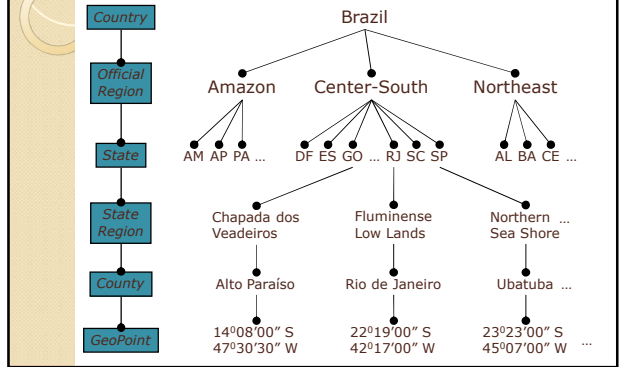
The Territory Dimension



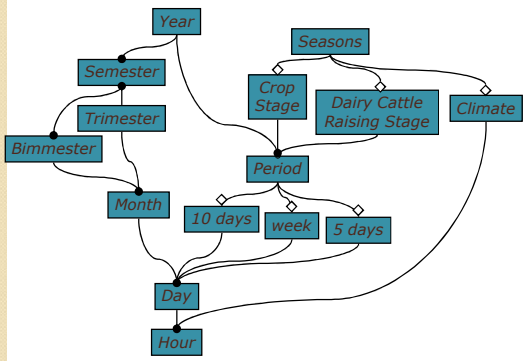
Instances of Territory (I)



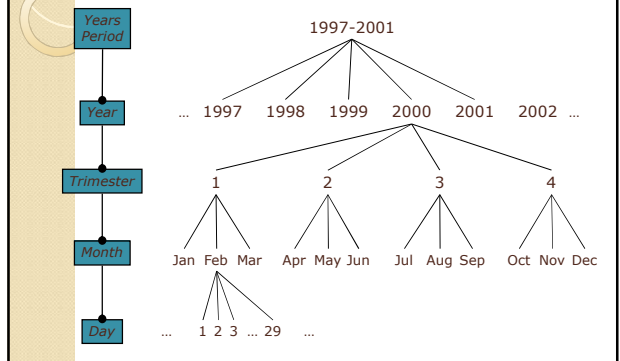
Instances of Territory (II)

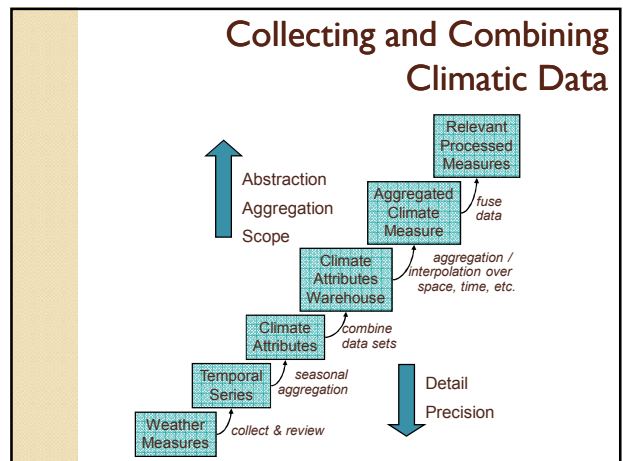
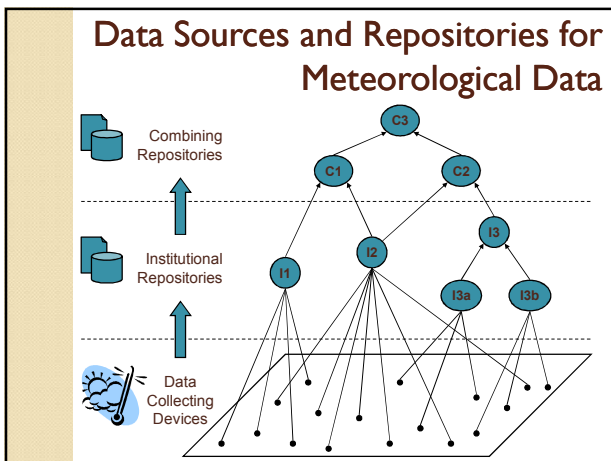
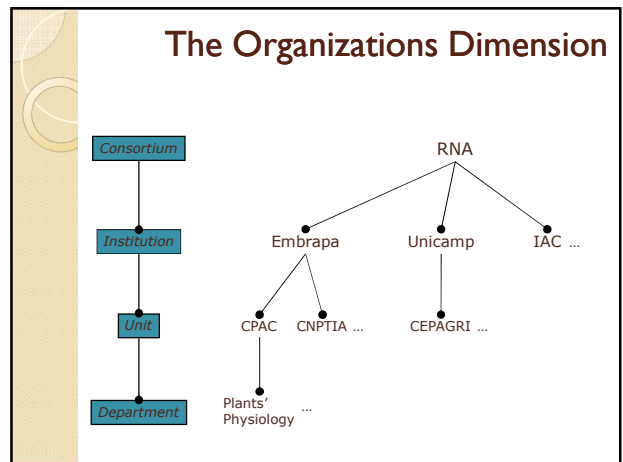
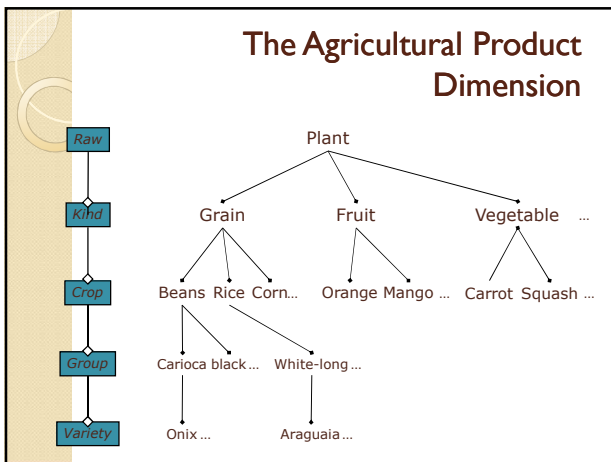
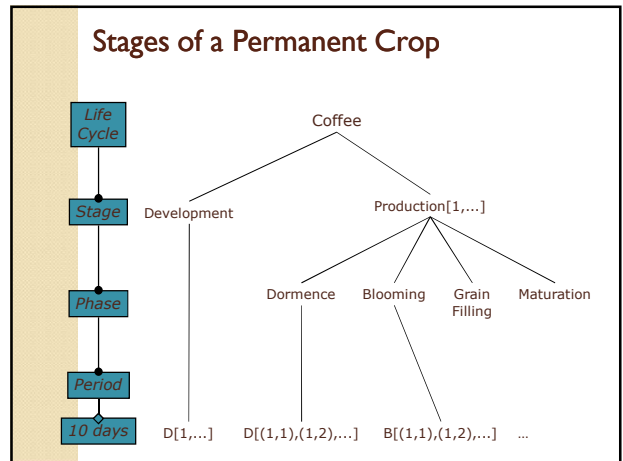
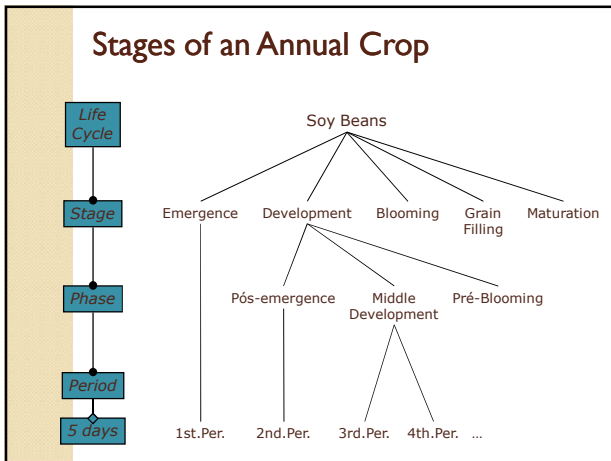


The Time Dimension

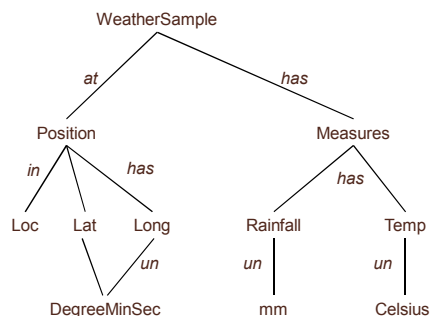


Calendar Time

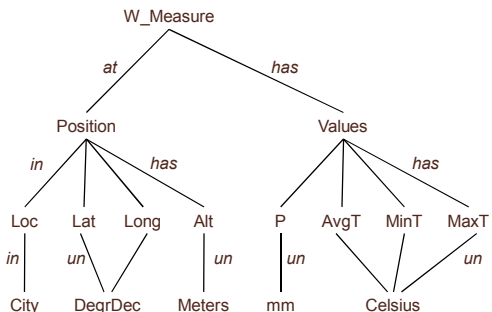




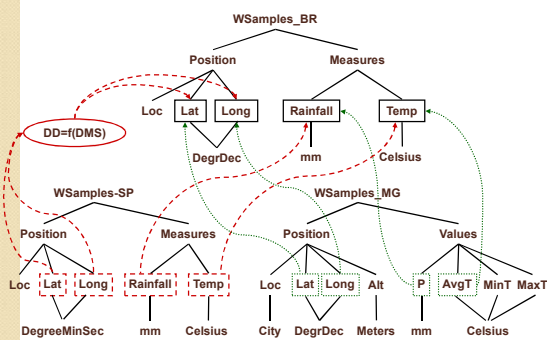
One Source of Weather Data



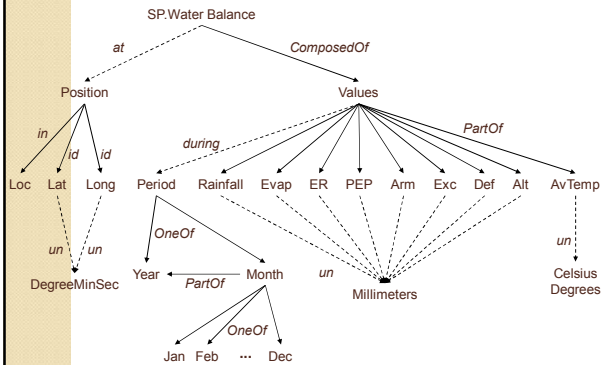
Another Source of Weather Data



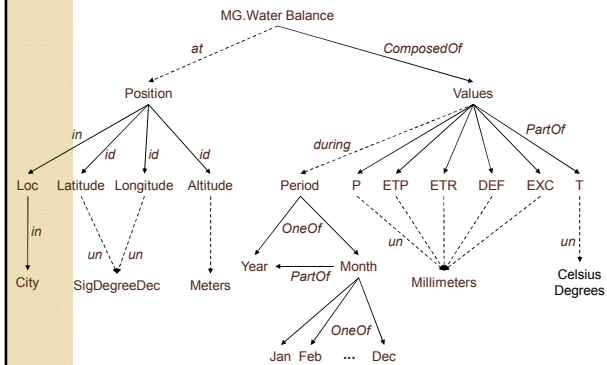
Data Integration



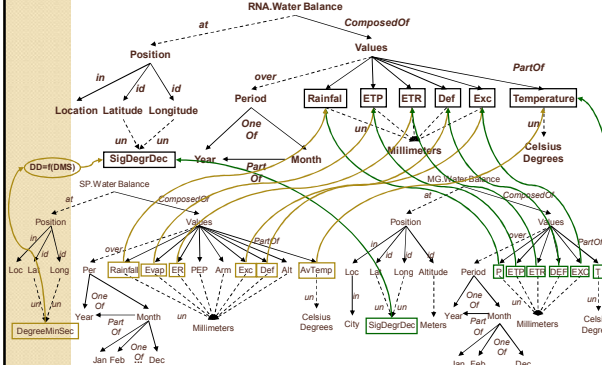
Water Balance São Paulo State



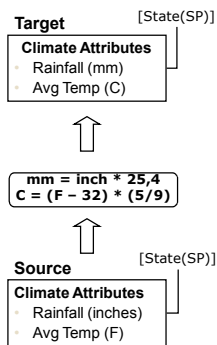
Water Balance Minas Gerais State



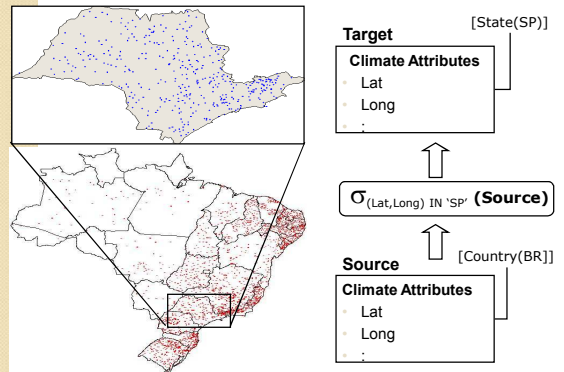
Water Balance Data Integration



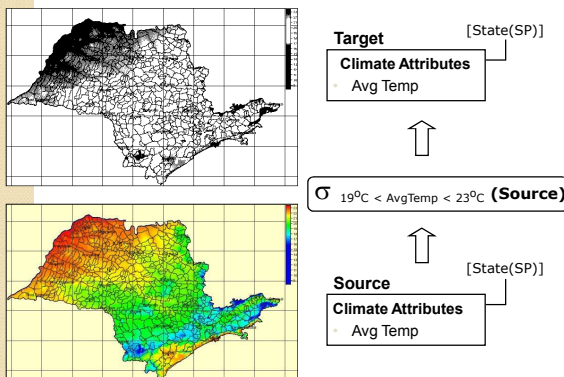
Converting Data Values



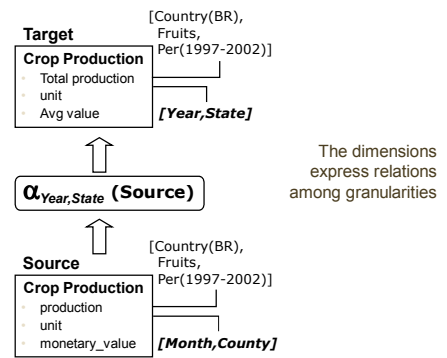
Selecting on a Dimension



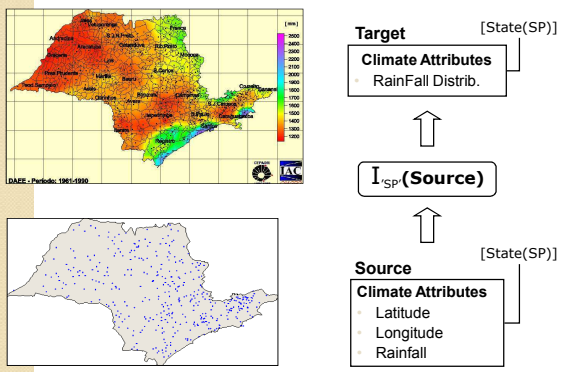
The Selection on a Data Value



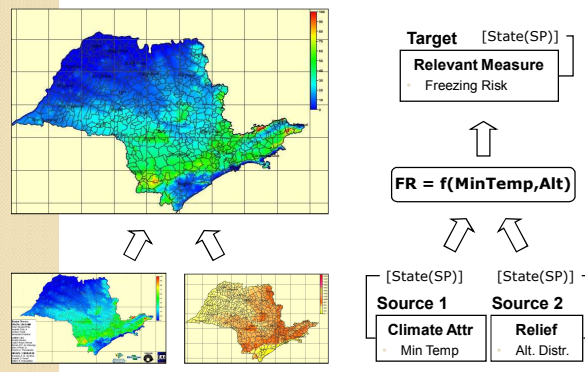
The Aggregation of Data Values



The Interpolation of Data Values



Combining Data Values



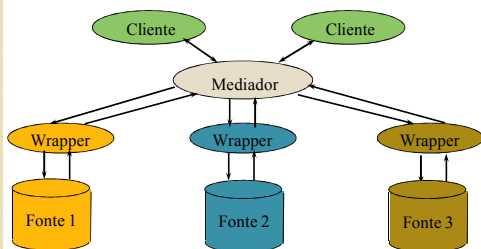
Tópicos

- **Conceitos básicos**
 - Sistemas de Suporte à Decisão
 - Data Warehouses (DWs)
 - OLTP vs. OLAP
 - Modelo Dimensional e operadores OLAP (*drill-down, roll-up, etc.*)
- **Planejando, projetando e implementando DWs**
 - Diretrizes gerais, fases do desenvolvimento e tarefas envolvidas
 - Estudo de Caso: DWs para a agricultura
- **Padrões e ferramentas para a implementação de DWs**
 - Principais componentes e padrões para DWs
 - Ferramentas abertas e proprietárias para DWs
- **Alguns temas de pesquisa atuais em DWs**
 - DWs com extensões espaciais e temporais
 - Semântica em DWs
- **Conclusões e referências para estudos adicionais**

Componentes típicos de um DW

- Ferramentas para Extração, Transformação e Carga (ETC) para facilitar a integração e assegurar a qualidade dos dados carregados (e.g., Keetle, GeoKettle)
- Mecanismos de interoperabilidade para fontes de dados heterogêneas (e.g., gateways como ODBC e JDBC, wrappers e mediadores)
- Sistema de gerenciamento de bancos de dados (com bom desempenho e segurança!)
- Catálogo de metadados (necessário para documentar e compreender os dados e suas transformações)
- Gerenciador de DataMarts
- Servidor OLAP (e.g. Mondrian)
- Gerador de interfaces cliente na Web (e.g. Jpivot)

Wrappers e mediadores



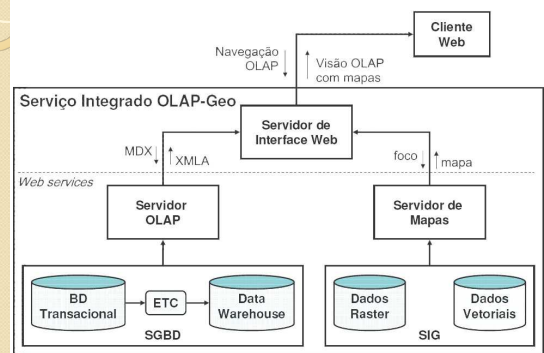
Implementação de OLAP

- **ROLAP**
 - Relacional
- **MOLAP**
 - Multidimensional
- **HOLAP**
 - Híbrido

Padrões abertos para módulos de DWs

- **XMLA (XML for Analysis)**: Padrão baseado em esquema XML para definição da estrutura de cubos, possibilidades de navegação (consultas) sobre o mesmo e comunicação entre cliente e servidor OLAP
- **MDX (MultiDimensional Expressions)**: Linguagem de consulta com sintaxe similar a SQL para especificação de consultas OLAP
- **JOLAP (Java OLAP)**: API para acesso a servidor OLAP a partir de aplicações Java

Arquitetura Geral



Especificação de cubo no Mondrian

```

<?xml version="1.0"?>
<Schema name="SampleData">
  <Cube name="Quadrant Analysis">
    <Table name="fato_desempenho">
      <Dimension name="Curso" foreignKey="id_curso">
        <Hierarchy hasAll="true" allMemberName="All Cursos" primaryKey="id_curso">
          <Table name="curso">
            <Level name="curso.centro" table="curso" column="centro" uniqueMembers="false">
            <Level name="curso.area" table="curso" column="area" uniqueMembers="false">
            <Level name="curso.nome" table="curso" column="nome" uniqueMembers="false">
          </Hierarchy>
        </Dimension>
        <Dimension name="disciplina" foreignKey="id_disciplina">
          <Hierarchy hasAll="true" allMemberName="All disciplinas" primaryKey="id_disciplina">
            <Table name="disciplina">
              <Level name="disciplina.areas_conhecimentos" table="disciplina" column="areas_conhecimentos" uniqueMembers="false">
              <Level name="disciplina.disciplina" table="disciplina" column="disciplina" uniqueMembers="false">
            </Hierarchy>
          </Dimension>
          <Dimension name="Candidato" foreignKey="id_candidato">
            <Hierarchy hasAll="true" allMemberName="All Candidatos" primaryKey="id_candidato">
              <Table name="candidato">
                <Level name="candidato.estado_endereco" table="candidato" column="estado_endereco" uniqueMembers="false">
                <Level name="candidato.cidade_endereco" table="candidato" column="cidade_endereco" uniqueMembers="false">
              </Hierarchy>
            </Dimension>
            <Dimension name="soma" column="soma_notas" aggregator="sum" datatype="String" formatString="###0">
              <Measure name="maxima" column="nota_max" aggregator="max" datatype="String" formatString="###0">
              <Measure name="minima" column="nota_min" aggregator="min" datatype="String" formatString="###0">
              <Measure name="num" column="num_candidatos" aggregator="sum" datatype="String" formatString="Standard">
            </Dimension>
          </Cube>
        </Schema>

```

Script com consulta MDX

```

...
<action-definition>
  <component-name>PivotViewComponent</component-name>
  <action-inputs>
    <cursoSelects type="string"/>
    <candidatosSelects type="string"/>
    <disciplinaSelects type="string"/>
  </action-inputs>
  <action-outputs> ... </action-outputs>
  <component-definition>
    <title>Drill Down to Pivot Table</title>
    <viewer>Pivot</viewer>
    <model>samples/analysis/query1.mondrian.xml</model>
    <jndi>dw_tcc</jndi>
  ...
  <query><![CDATA[with member [Measures].[Media notas] as '([Measures].[soma] / [Measures].[num])', format_string = IIF((( [Measures].[soma] / [Measures].[num])) > 4.0), '###.0|style="green"', IIF((( [Measures].[soma] / [Measures].[num])) < 2.0), '###.0|style="red"', '###.0')|select NON EMPTY ([Measures].[maxima], [Measures].[Media notas]) ON COLUMNS, NON EMPTY Crossjoin({ [cursoSelects] }, Crossjoin({ [candidatosSelects] }, { [disciplinaSelects] } ) ) ON ROWS from [Quadrant Analysis]]]></query>
  ...
</action-definition>
...

```

Comparativo de ferramentas livres (Giovani Caruso McDonald e João Rubik, UFSC, 2007)

Critérios agrupados segundo classes de ferramentas

- Todas as ferramentas
- Ferramentas de ETC
- Servidores OLAP
- Clientes OLAP
- Suites de ferramentas para desenvolvimento de DWs

Critérios Gerais

- Sistema Operacional
- Licença
- Usabilidade
- Linguagem de Desenvolvimento

Ferramentas de ETC

- Implementação (ROLAP, MOLAP, etc)
- Fontes de dados
- Limpeza de dados
- Abrangência dentro do processo

Servidores OLAP

- SGBDs suportados
- Metodo de armazenamento
- Protocolo de conexão
- Linguagem de consulta

Clientes OLAP

- Linguagem de consulta
- Forma de conexão com o servidor
- Gerência de análises

Suites de ferramentas

- Componentes utilizados e compatíveis
- Administração

Licenças

- GNU General Public License (GPL)
- GNU Library General Public License (LGPL)
- Common Public License (CPL)
- Mozilla Public License
- Apache License
- BSD License

Ferramentas livres analisadas

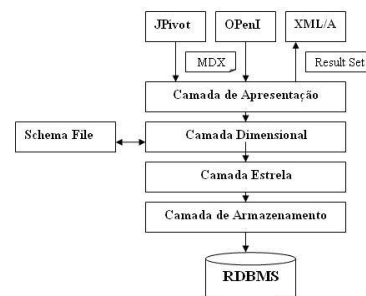
- Mondrian (servidor OLAP)
- Jpivot (geração de interfaces OLAP)
- OpenI (suite)
- SpagoBI (suite)
- PentahoBI (suite)
- BEE Project (suite)

Mondrian

- Busca os dados de um sistema de gerenciamento de banco relacional e apresentando o resultado no formato multidimensional
- Executa consultas a ele submetidas na linguagem MDX.
- O formato de saída é processado por seus clientes OLAP

Mondrian

- Arquitetura:



Mondrian

- Instalação e configuração
 - Disponível no *SourceForge*.
 - Definir o esquema do cubo multidimensional
 - O esquema é descrito por um documento XML

JPivot

- É uma biblioteca JSP para a interação do usuário com servidores OLAP via Web
- O JPivot permite aos usuários submeter consultas a servidores OLAP e explorar a resposta obtida através de uma interface gráfica

JPivot

- Arquitetura
 - Suporta diversos servidores OLAP
 - Tipos de conexão com servidores:
 - Acesso através de XMLA
 - Acesso através de um Driver JDBC
 - Utiliza o pacote WCF (*Web Component Framework*) para a construção da Interface

JPivot

Curso	disciplinaprova	Candidato	Measures	
			maximo	num
*All cursos	*All disciplinas	*All Candidatos	10.00	318,468
		-Behia	9.49	324
		-ILHEUS	8.84	12
		-JEQUE	8.75	12
		-PORTO SEGURO	7.30	12
		-SALVADOR	9.49	216
		*Goiás	9.80	1,248
		*Mato Grosso	10.00	1,548
		*Mato Grosso do Sul	9.80	2,256
		*Minas Gerais	9.83	1,704
		*Paraná	10.00	17,208
		*Rio de Janeiro	9.50	552
		*Rio Grande do Sul	9.67	10,752
		*Santa Catarina	10.00	250,128
		*Sao Paulo	10.00	29,484

Openl

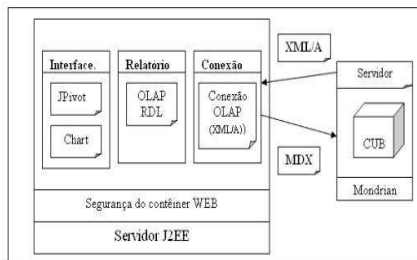
- Openl é suite de BI (Business Intelligence) Open Source baseada em tecnologia J2EE para análise OLAP
- Compatível com servidores OLAP que utilizam o protocolo XMLA
- Propósito de ser uma ferramenta fácil e intuitiva para executar operações OLAP.

Openl

- Arquitetura
 - O componente de conexão
 - O componente de Relatório
 - Componente de interface
 - O Componente de Segurança

Openl

- Arquitetura



Openl

- Instalação e Configuração

- Aplicação Web que funciona em qualquer servidor de aplicação J2EE
- Segurança do Openl é baseada na segurança do servidor de aplicação J2EE
- Utiliza o conceito de projetos para definir uma coleção específica de relatórios OLAP

Openl

- Criação de análise

A imagem mostra a interface 'New Analysis' do Openl. Ela contém campos para 'Datasource' (mondrian), 'Catalog' (Vestibular) e 'Cube' (Desempenho). Abaixo, há uma seção 'Configure analysis dimensions and measures' com tabelas para 'Columns', 'Rows' e 'Filter'. O filtro selecionado é 'disciplinaprova'. Botões 'OK' e 'Cancel' estão na base.

Pentaho

- Objetivo de fornecer soluções para os problemas em BI
- Arquitetura orientada a serviços e centrada em processos
- Engloba componentes responsáveis por relatórios, análises, mineração de dados, ETC, entre outros

Pentaho

- Arquitetura:

- Componentes maduros
- Projetos consolidados em código aberto
- Facilmente substituídos por outros produtos
- Padrão de servidor web J2EE
- Repositórios de dados é externo a plataforma Pentaho

Pentaho

- Componentes:

- Servidor OLAP Mondrian
- Front-End JPivot Analysis
- Shark e JaWEWorkflow
- Kettle EII e ETL
- Servidor de aplicação JBoss, Hibernate e Portal
- Weka Data Mining
- EclipseWorkbench e BIRT reporting components
- JOSSO single sign-on and LDAP integration
- Mozilla Rhino Javascript Processor

Pentaho

- Visualização através do JPivot

Drill Down to Pivot Table

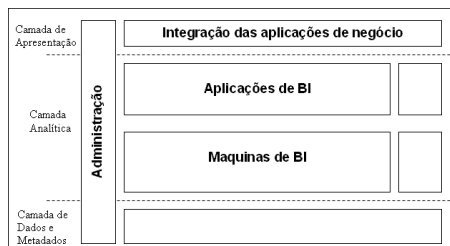
Region	Department	Positions	Measures			
			Actual	Budget	Variance	Variance Percent
-All Regions	+All Departments	+All Positions	143.639.982,00	143.199.389,00	-440.593,00	-,31%
Central	+All Departments	+All Positions	37.893.162,00	38.397.600,00	504.438,00	1,31%
Eastern	+All Departments	+All Positions	35.248.940,00	35.487.861,00	238.921,00	,67%
Southern	+All Departments	+All Positions	35.248.940,00	34.903.861,00	-445.079,00	-1,26%
Western	+All Departments	+All Positions	35.248.940,00	34.510.067,00	-738.873,00	-2,14%

SpagoBI

- Suite que engloba todos as etapas do processo de BI. Desde o ETC chegando até as ferramentas de Data Mining
- Ferramentas adicionadas como componentes dentro de um portal

SpagoBI

- Arquitetura



SpagoBI

- Instalação
 - Arquivo executável instala o Portal e os componentes necessários para o funcionamento da plataforma.

Spago BI

A imagem mostra a interface do usuário do Spago BI. No topo, há o menu 'Home' com opções como 'President Desc', 'Navigation', 'Reports', 'OLAP', 'ONE', 'User', 'BreakOut' e 'Data Mining'. Abaixo, há uma barra de ferramentas e uma área de trabalho com painéis para 'Columns', 'Rows', 'Filters' e 'MDX Query Editor'. O editor de consultas MDX contém o seguinte código:

```
select NON EMPTY ([Measures].[Store Sales], [Measures].[Store Cost], [Measures].[Unit Sales]) ON COLUMNS,
NON EMPTY Order(Union([Region].[All Regions],[Central West],[Region].[All Regions],[South West],[Los Angeles],[Region].[All Regions],[South West],[Los Angeles],[Beverly Hills],[Region].[All
```

BEE Project

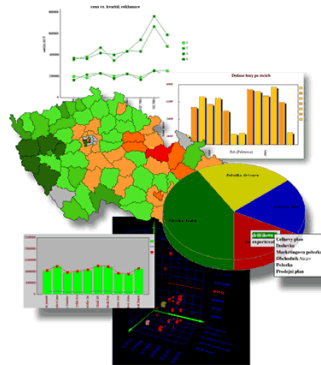
- Suite que reúne uma ferramenta ETC, um servidor OLAP e duas opções de cliente, um desktop e um web



BEE Project

- A instalação do BEE Project é a mais complexa entre as ferramentas analisadas, e exige, entre outros, a compilação do código fonte PERL. Além disso a documentação é muito deficiente o que dificulta e toma muito tempo. Por esses motivos esse trabalho analisou o BEE project apenas através da documentação disponível e a instalação não foi realizada.

BEE Project



Comparativo de ferramentas livres (Giovani Caruso McDonald e João Rubik, UFSC)

Ferramenta	Mondrian	Openl	Jpivot	Pentaho	Bee Project	Spago BI
Categoria	Servidor	Cliente	Cliente	Suite	Suite	Suite
Sistema Operacional	Várias (JVM)	Várias (JVM)	Várias (JVM)	Várias (JVM)	POSIX (Linux/BSD/ UNIX similares), Linux	Várias (JVM)
Ling. de desenv.	Java	Java	Java	Java	C, Perl	Java
Licença	CPL	OPL	CPL	PPL	GPL	GPL
Usabilidade	Pouco Amigável	Muito Amigável	Amigável	Muito Amigável	Pouco Amigável	Amigável
SGBD suportado	JDBC (todos)	Não se Aplica	Não se Aplica	JDBC (todos)	Perl DBI/DBD, MySQL, Oracle, PostgreSQL (pgsql)	JDBC (todos)

Comparativo de ferramentas livres (Giovani Caruso McDonald e João Rubik, UFSC)

	Mondrian	JPivot	Openl	Pentaho	BEE Project	SpagoBI
Método de Armazenamento	Rolap	Não se Aplica	Não se Aplica	Não se Aplica	Não se Aplica	Não se Aplica
Formas de Conexão Disponíveis	XMLA, JDBC	Não se Aplica	Não se Aplica	Não se Aplica	Não se Aplica	Não se Aplica
Linguagem de Consulta	MDX	MDX	MDX	Não se Aplica	Não se Aplica	Não se Aplica
Formas de Conexão com Servidor	Não se Aplica	XML/A, JDBC	XML/A, JDBC	Não se Aplica	Não se Aplica	Não se Aplica
Gerencia de análises	Não se Aplica	Sim	Não	Sim	Sim	Sim
Componente Servidor	Não se aplica	Não se aplica	Não se aplica	Mondrian		Mondrian

Ferramentas proprietárias

- Microsoft (SQL Server + Analysis Services)
- Oracle
- IBM
- Informix
- Cartelon
- NCR
- Red Brick
- Sybase
- SAS
- Microstrategy Corporation
- Pentaho versão comercial
- :

Tópicos

- **Conceitos básicos**
 - Sistemas de Suporte à Decisão
 - Data Warehouses (DWs)
 - OLTP vs. OLAP
 - Modelo Dimensional e operadores OLAP (*drill-down, roll-up, etc.*)
- **Planejando, projetando e implementando DWs**
 - Diretrizes gerais, fases do desenvolvimento e tarefas envolvidas
 - Estudo de Caso: DWs para a agricultura
- **Padrões e ferramentas para a implementação de DWs**
 - Principais componentes e padrões para DWs
 - Ferramentas abertas e proprietárias para DWs
- **Alguns temas de pesquisa atuais em DWs**
 - DWs com extensões espaciais e temporais
 - Semântica em DWs
- **Conclusões e referências para estudos adicionais**

DWs espaciais

Problemas em aberto:

- Integração do modelo dimensional com algum modelo espacial:
 - Modelagem
 - Operadores
 - Implementação de sistemas integrados
- Integração de dados em aplicações
- Geração de datamarts

Operadores e funções de agregação em DW

• Operadores:

- Agregação/desagregação: Roll-up, Drill-down, group by
- Seleção/projeção: Pivot, Slice e Dice

• Funções de agregação:

- Distributivas: contagem, mínimo, máximo, soma
- Algébricas: média, desvio padrão
- Holísticas: mediana, maior frequência, rank

Operadores em SIG

• Operadores:

• Egenhofer propôs:

- “disjunto de ...”
- “encontram...”
- “é igual a ...”
- “dentro de ...”
- “contém ...”
- “cobre ...”
- “é coberto por ...”
- “sobreposição”

Classificação de Operadores em SIG

• De acordo com o número de operandos:

- Unários
- Binários
- N-ários

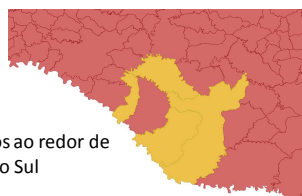
• De acordo com o tipo do dado retornado:

- Booleano
- Escalar
- Espacial

Operadores em SIG

• Unário com resultado espacial:

Ex.: Municípios ao redor de Campo Belo do Sul

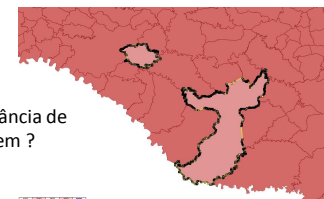


Operadores em SIG

• Binário com resultado escalar:

- Distância
- Área

Ex.: Qual a distância de Lages até Vargem ?
Resp. 64 km



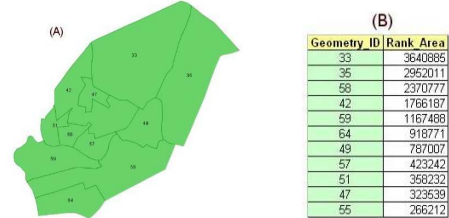
Funções de agregação em DWs espaciais (Joel Silva, UFPE, 2008)

- Proposta de nova classificação
- Funções de agregação classificadas de acordo com o tipo da função e com o tipo de retorno:

Distributivas, Algébricas e Holísticas
X
Escalar e Espacial

Funções de agregação em DWs espaciais (Joel Silva, UFPE, 2008)

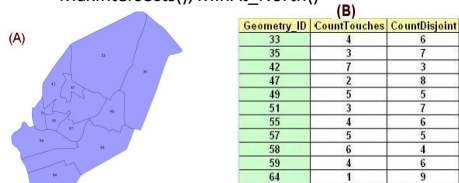
- Holística escalar:
 - RankArea(), RankPerimeter(), MedianArea(), ModeArea()



Ex.: A partir de uma figura (A), indicar as áreas em um ranking (B)

Funções de agregação em DWs espaciais (Joel Silva, UFPE, 2008)

- Distributiva escalar:
 - CountTouches(), CountAt_North_of(), MaxIntersects(), MinAt_North()

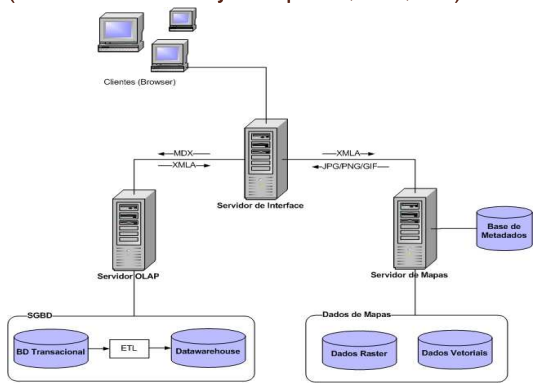


Ex.: A partir de uma figura (A), indicar quantos objetos tocam e quantos estão disjuntos (B)

Funções de agregação em DWs espaciais (Joel Silva, UFPE, 2008)

- Distributiva espacial:
 - SumTouches(), SumAt_North_of(), SumAt_North_West_of(), SumDisjoint()
- Algébrica escalar:
 - AvgArea(), StdvPerimeter(), MaxNArea()
- Algébrica espacial:
 - MaxNat_North_of(), MaxNat_South_of()

Integrando Ferramentas OLAP e SIG (Carlos Eduardo Costa e José Felipe Neis, UFSC, 2007)

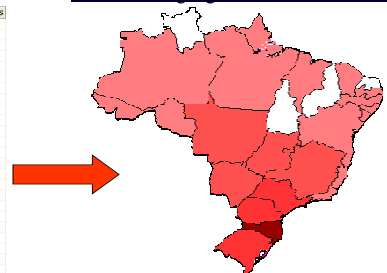


Visualizando dados em mapas (Miguel Soares e Nuno Santos, ISCTE-Lisboa e UFSC, 2008)

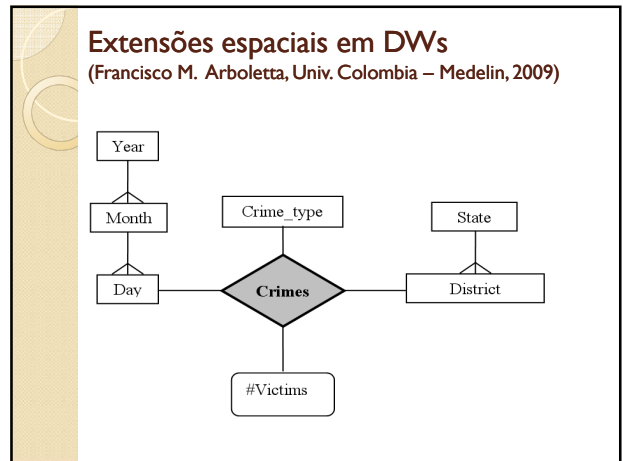
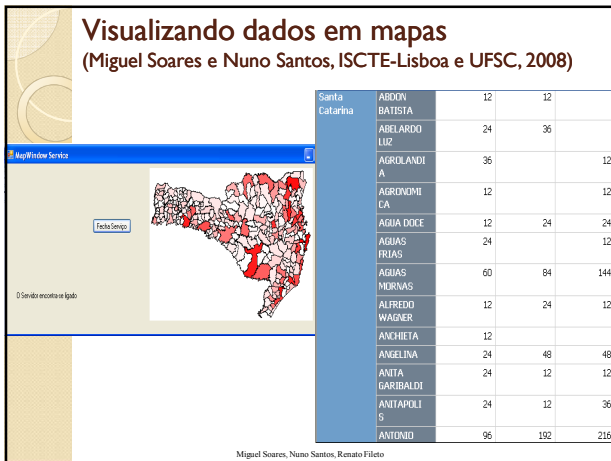
Vista tabular

Estado	Endereco	Num Candidatos
Acre		48
Alagoas		168
Amapá		24
Amazonas		48
Bahia		648
Ceará		72
Distrito Federal		4800
Espírito Santo		360
Goiás		2496
Maranhão		24
Mato Grosso		3096
Mato Grosso do Sul		4512
Minas Gerais		3408
Pará		120
Paraná		24
Paraná		34416
Pernambuco		48
Rio de Janeiro		1104
Rio Grande do Sul		21504
Rorônia		364
Santa Catarina		500256
São Paulo		58968
Sergipe		264
Tocantins		120
Grand Total		636912

Vista geográfica



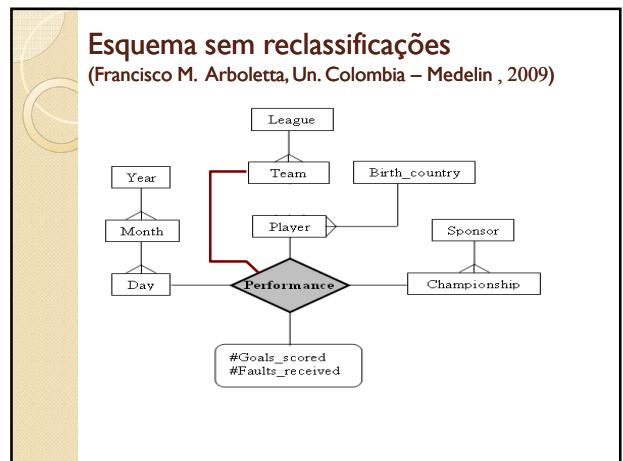
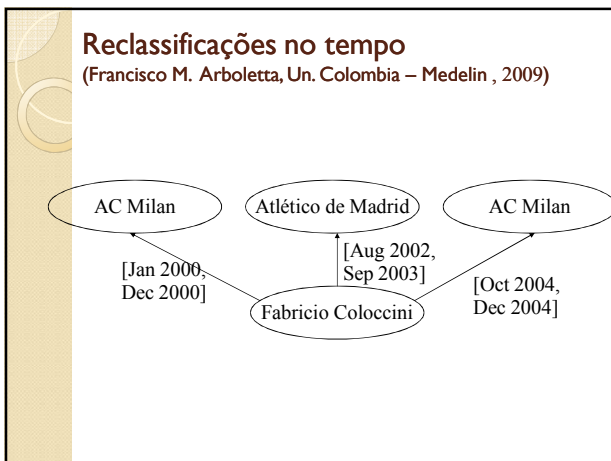
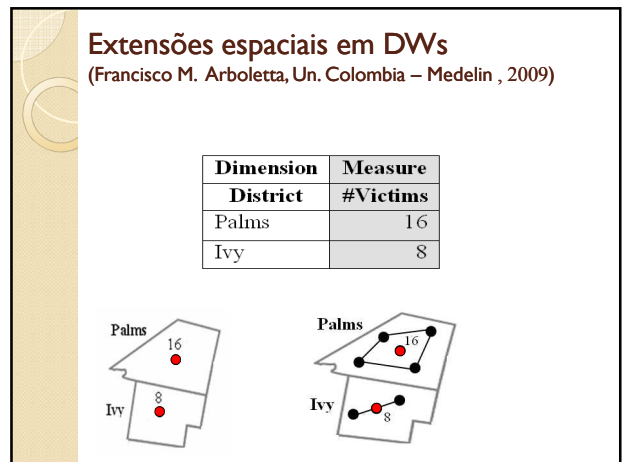
Miguel Soares, Nuno Santos, Renato Fíleto



Extensões espaciais em DWs

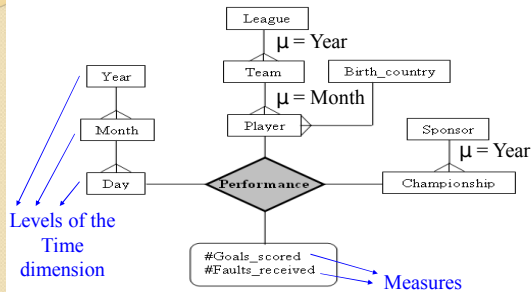
(Francisco M. Arboletta, Un. Colombia – Medellin, 2009)

Dimensions			Measures	
District	Crime_type	Date	#Victims	Crime_points
Palms	Robbery	Jan 5/07	6	{P ₁ }
Palms	Robbery	Jan 6/07	9	{P ₂ , P ₃ }
Palms	Murder	Jan 5/07	1	{P ₄ }
Ivy	Robbery	Jan 5/07	3	{P ₅ }
Ivy	Murder	Jan 6/07	5	{P ₆ }



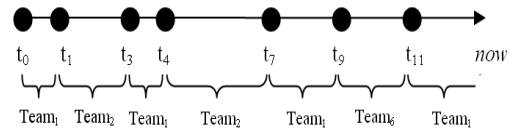
Reclassificações com granularidade

(Francisco M. Arboletta, Un. Colombia – Medellin , 2009)



Histórico de Reclassificações

(Francisco M. Arboletta, Un. Colombia – Medellin , 2009)



Evolution of the assignments of player P_1 to teams

Consultas Temporada (Season Queries)

(Francisco M. Arboletta, Un. Colombia – Medellin , 2009)

- What was the total number of goals scored by P_1 in his first and last seasons in $Team_1$? In his first three seasons in $Team_1$? In his last two seasons in $Team_1$?
- Which was the season (dates, number of season, and team), where the total number of goals scored by P_1 was the highest?
- What was the total and average number of goals scored by each player in each season for each team?

Tópicos

- **Conceitos básicos**
 - Sistemas de Suporte à Decisão
 - Data Warehouses (DWs)
 - OLTP vs. OLAP
 - Modelo Dimensional e operadores OLAP (*drill-down, roll-up, etc.*)
- **Planejando, projetando e implementando DWs**
 - Diretrizes gerais, fases do desenvolvimento e tarefas envolvidas
 - Estudo de Caso: DWs para a agricultura
- **Padrões e ferramentas para a implementação de DWs**
 - Principais componentes e padrões para DWs
 - Ferramentas abertas e proprietárias para DWs
- **Alguns temas de pesquisa atuais em DWs**
 - DWs com extensões espaciais e temporais
 - Semântica em DWs
- **Conclusões e referências**

Conclusões

- DWs permitem a integração de dados e a execução de análises detalhadas e dinâmicas (OLAP) da informação, na forma de tabelas, gráficos e mapas para apoio à tomada de decisão
- A disponibilidade de ferramentas livres ou de baixo custo para a implementação de DWs abre oportunidades para a aplicação desta tecnologia em pequenos e médios empreendimentos
- Aconselha-se o desenvolvimento de DWs de forma gradual, ao invés de tentar alcançar todos os objetivos de uma única vez

Conclusões (II)

- Padrões de sistemas abertos possibilitam a interoperabilidade de componentes na Web (SGBDs, servidores OLAP, servidores e interfaces para tabelas, gráficos e mapas, etc.)
- Diversas aplicações requerem tratamento especial das dimensões espaço e tempo, gerando desafios na integração com GIS e uso de técnicas de BDs temporais
- Semântica pode auxiliar na integração de dados e projeto de aplicações de DWs (e.g., geração de DMs)

Links úteis

- **Suítas de ferramentas para DWs**
 - <http://www.pentaho.org/>
 - <http://spagobi.objectweb.org/>
 - <http://www.microsoft.com/sqlserver/2008/en/us/default.aspx>
 - <http://www.microsoft.com/sqlserver/2008/en/us/wp-sql-2008-analysis-services.aspx>
 - <http://openi.sourceforge.net>

Links úteis (II)

- **Outras ferramentas** (algumas embutidas em suítas)
 - <http://mondrian.sourceforge.net>
 - <http://pivot.sourceforge.net/>
 - <http://sourceforge.net/projects/bee/> (MySQL)
 - <http://www.r-project.org/>
 - <http://www.eclipse.org/birt/phoenix/>
 - <http://rubik.sourceforge.net/jrubik/intro.html>
 - <http://www.jfree.org/jfreechart/> (gráficos)
 - <http://jasperreports.sourceforge.net/>
 - <http://www.palo.net/>
 - <http://kettle.pentaho.org/>

Links úteis (III)

- **Padrões para conexão com servidores OLAP**
- MDX**
- <http://www.xmlforanalysis.com/mdx.htm>
 - <http://www.databasejournal.com/features/mssql/article.php/1495511>
- XMLA**
- <http://www.xmla.org>
 - <http://www.xmlforanalysis.com/>
 - <http://msdn2.microsoft.com/en-us/library/ms187178.aspx>

Referências bibliográficas

- Inmon, W.H. *Tech Topic: What is a Data Warehouse?* Prism Solutions. Volume 1. 1995.
- Kimball. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*, Wiley, 2002.
- Kimball, et al. *The Data Warehouse Lifecycle Toolkit*, Wiley, 1998.
- Kimball and Caserta. *The Data Warehouse ETL Toolkit*, Wiley, 2004.
- Bernard Lupin. *Try OLAP! 2007*.
- Daniel Lemire. *Data Warehousing and OLAP - A Research-Oriented Bibliography*, 2007.
- Gray, Jim; Bosworth, Adam; Layman, Andrew; Priahesh, Hamid. Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals. *Proceedings of the 12th International Conference on Data Engineering*. IEEE., 1995. 152-159.
- Gartner Reveals Five Business Intelligence Predictions for 2009 and Beyond. <http://www.gartner.com/it/page.jsp?id=856714>.