enabler
Solutions for Retailing

Data Warehousing

Engª Informática e Computação

FEUP

27 de Maio de 2002

Agenda

- *Perspectiva de Negócio*
- *Perspectiva Técnica*
- *Perspectiva do Projecto*
- *Razões para construir um DW*
- *Modelação de Dados*
- *Exploração de Dados*
- *DW como processo*
- *Leitura recomendada*

Perspectiva de Negócio

Agenda

Negócio

Técnica

Projecto

- ✓ Sistema Analítico vs Sistemas Operacional

  Análise e Acção - Exemplo de aplicações:

  "Balanced Scorecard" - orçamento

  Nível de serviço de Fornecedores - rupturas / negociações

  Rentabilidade de Espaço - comparação entre lojas / ...

  Preços de Venda, Stocks - Excepção e acção

- ✓ Sistema analítico

  Possibilidade de cruzar informação de diversas fontes

  Facilidade de acesso aos dados e sua agregação - query Wizzard

  Extracção de dados para Excel -pivot tables

Perspectiva Técnica

Agenda

Negócio

Técnica

Projecto

- ✓ Repositório de dados
  - *Modelação de dados - Redundância de dados vs Normalização*
  - *Modelo em Estrela - Factos e Dimensões*
    - Performance - bons tempos de resposta ao utilizador
    - Cruzamento de dados provenientes de diversas fontes
- ✓ Exploração de Dados
  - *OLAP, ROLAP, MOLAP, Query Ad-hoc, Reporting*

Perspectiva do Projecto

Agenda

Negócio

Técnica

Projecto

- ✓ Deve ter inicio e fim, conhecidos
- ✓ Perspectivas de resultados, conhecidos
- ✓ Comunicação

  A organização deve saber com o que contar, e quando.

  Todos devem saber o que lhes compete no projecto, e quando

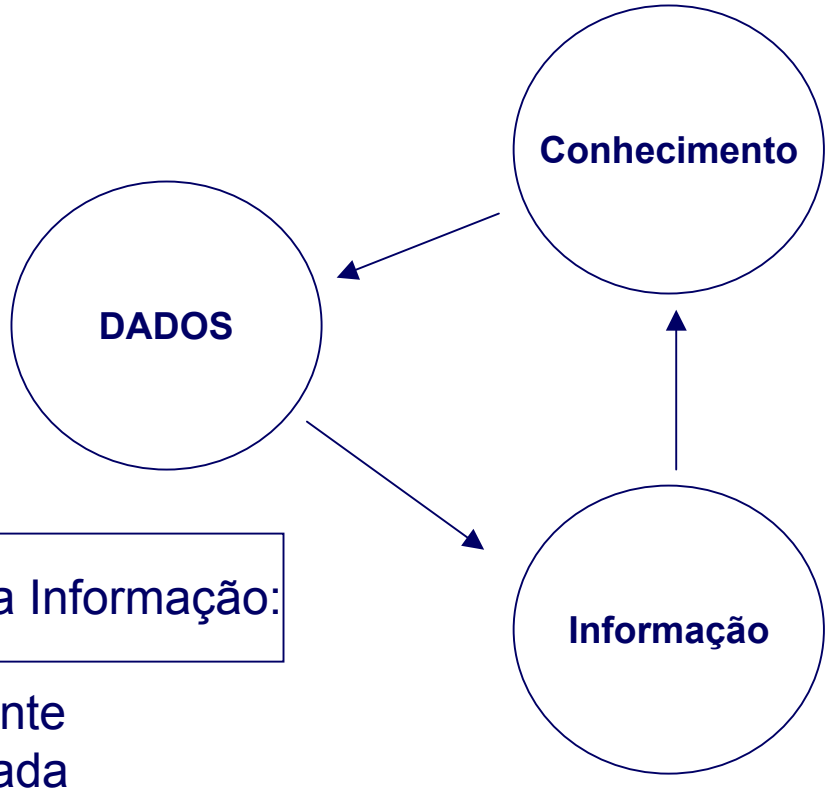
- ✓ Envolvimento do negócio

  Fases de Análise, Desenho, Testes de aceitação e Implementação

- ✓ Riscos: qualidade de dados

Razões para construir um DW

- ✓ Dados vs. Informação



Características da Informação:

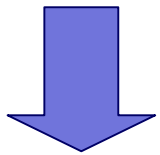
- Relevante
- Atempada
- Precisa
- Orientada à Acção

Razões para construir um DW

✓ Dispersão das fontes de dados

BANCA	**RETALHO**	**TELECOMUNICAÇÕES**
1. B.O.	**1.** F.O.	**1.** Billing
2. F.O.	**2.** ERP	**2.** Customer Care
3. Call-Center	**3.** Entreposto	**3.** Engenharia (rede)
4. Gestão de Produtos (específicos)	**4.** Finantials	**4.** Finantials

- Multiplicidade de Aplicações
- Multiplicidade de Interfaces
- Multiplicidade de Sistemas

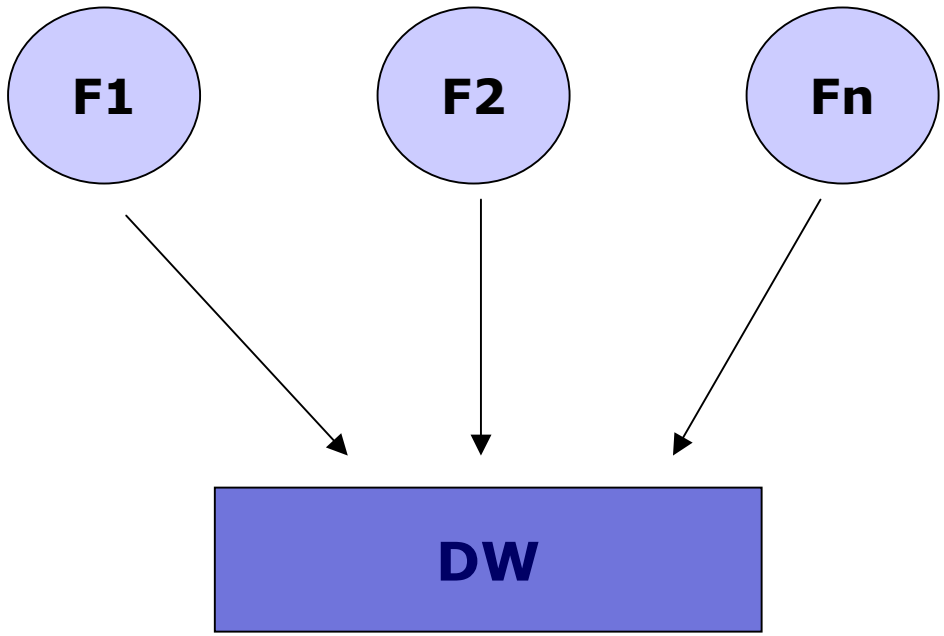


- Visibilidade dificultada
- Orientação operacional (não analítica) das fontes de dados

Razões para construir um DW

- ✓ Necessidades de Convergência + Navegação

  Visibilidade e Navegação: num único sistema

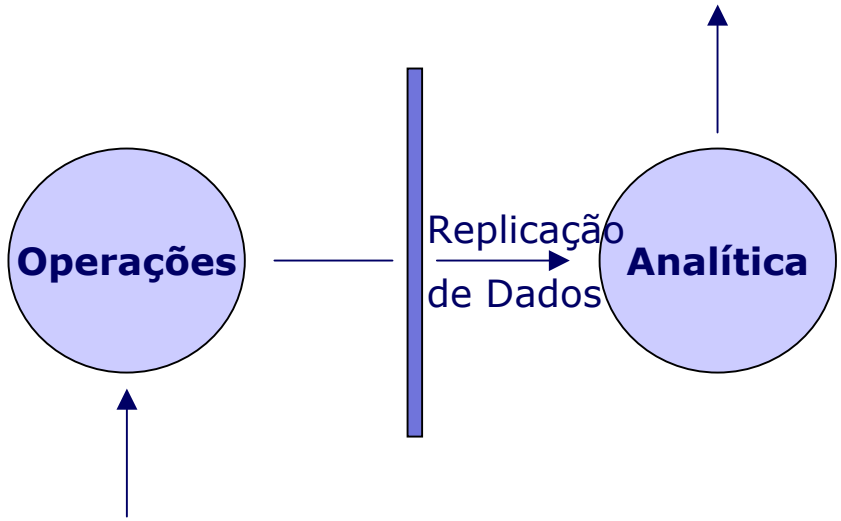


Razões para construir um DW

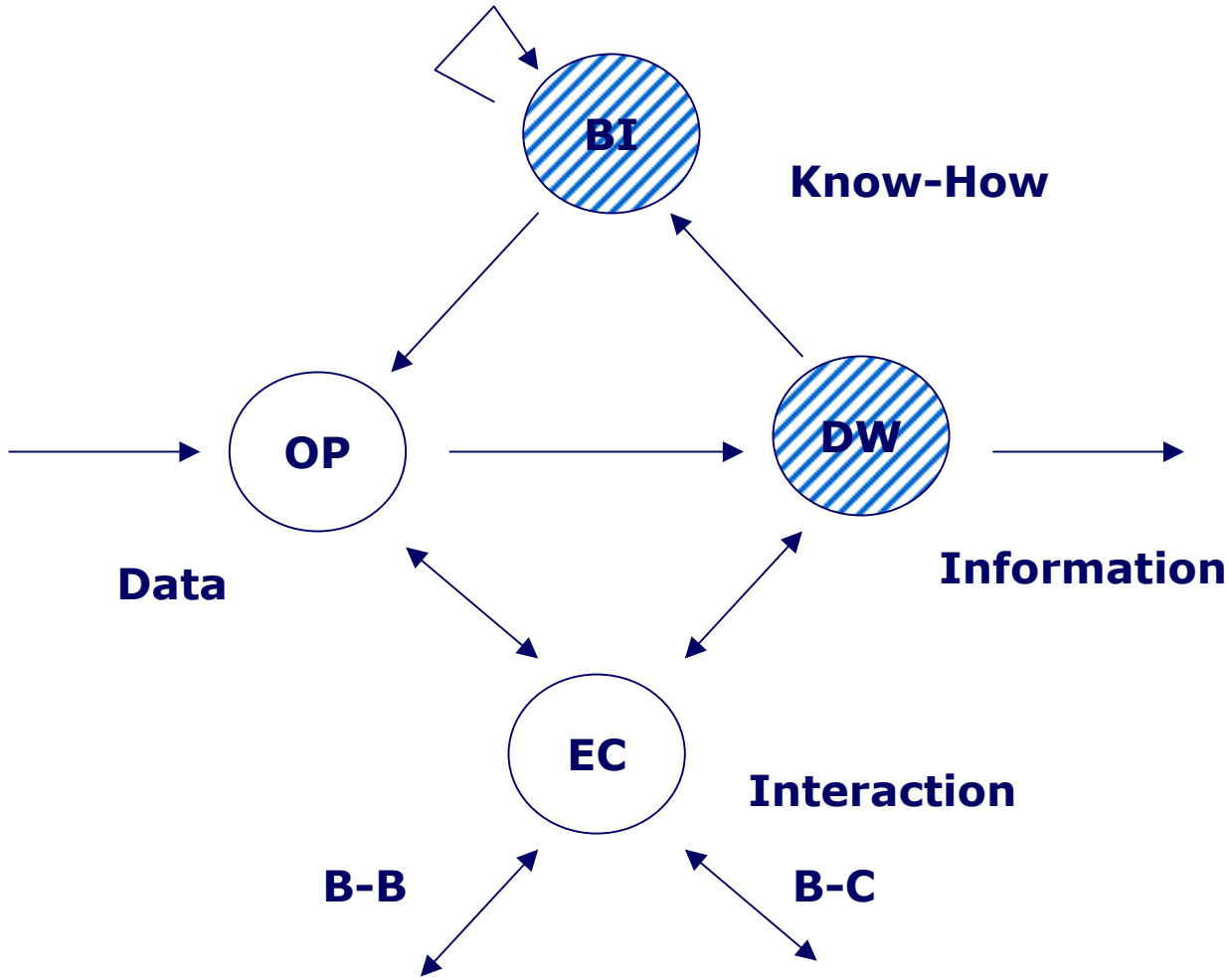
- ✓ Separação Operações / Analítica

  Isolar o impacto das explorações analíticas das operações

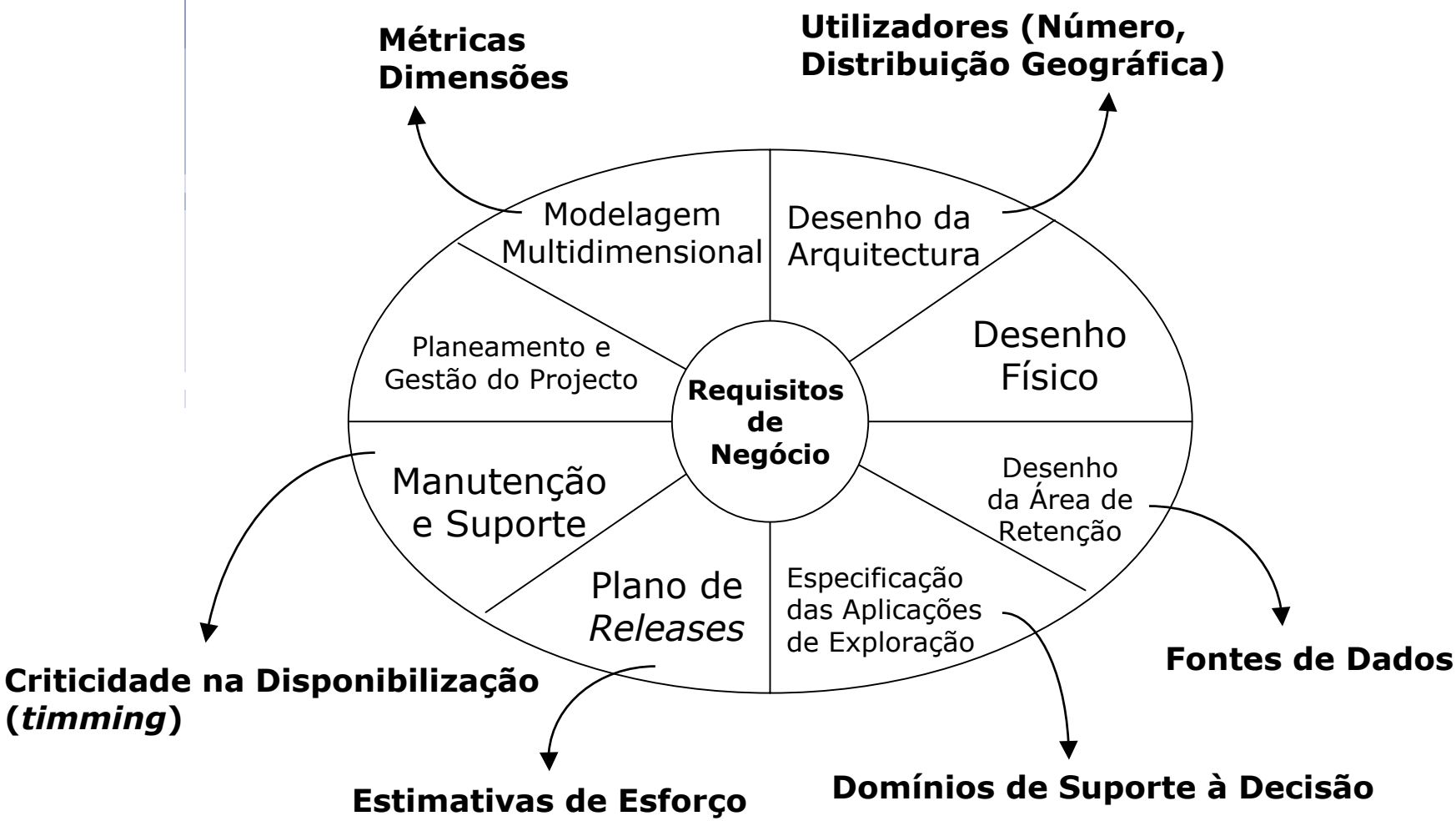
  Nasce a janela nocturna do batch de replicação



Enquadramento do DW numa Meta-Arquitectura Organizacional

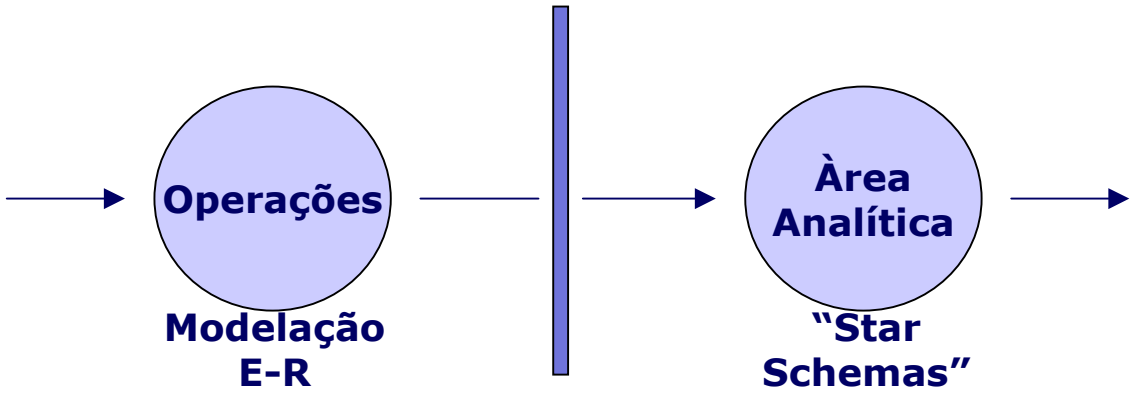


Levantamento de Requisitos



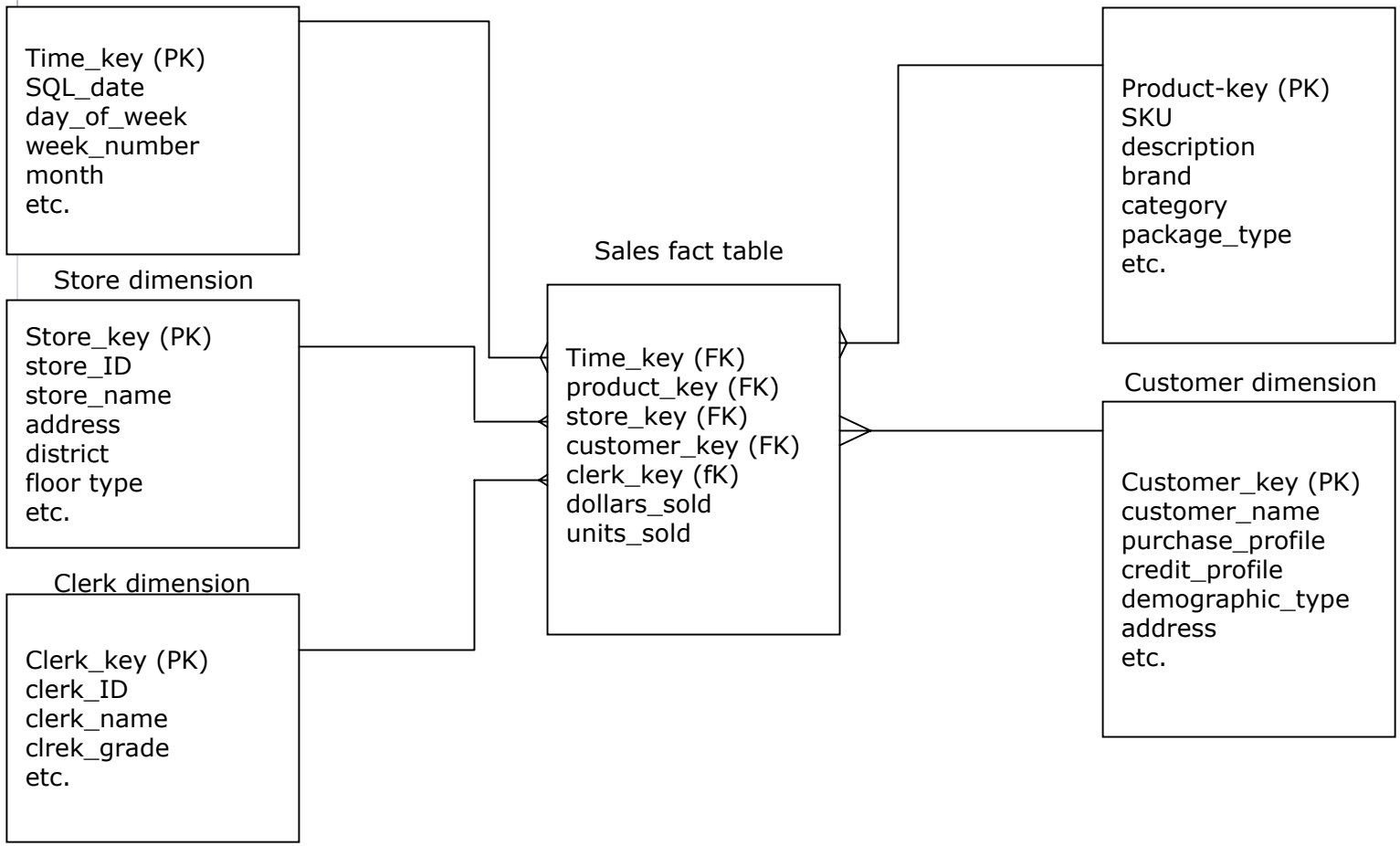
Modelação de Dados

- ✓ Nas Operações: 3ª Fórmula normal é Lei! (CODD)
- ✓ Na área analítica: Emergência do conceito de "Star Schema" (Kimball)



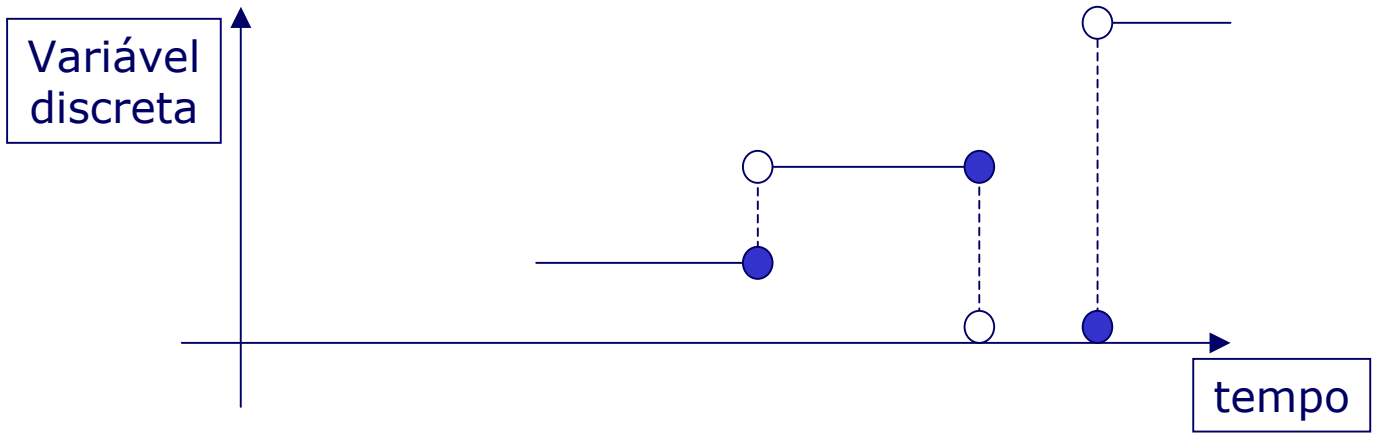
Modelação de Dados

- ✓ "Star schema"



Modelação de Dados

✓ Modelação do Tempo



- *As transicções de estado não estão normalmente arquivadas (história das alterações) nas bases operacionais; estas só guardam o último valor da variável*
- *Para efeitos de DW, este aspecto é crucial (Análise baseada em históricos, análise de evolução e tendências)*

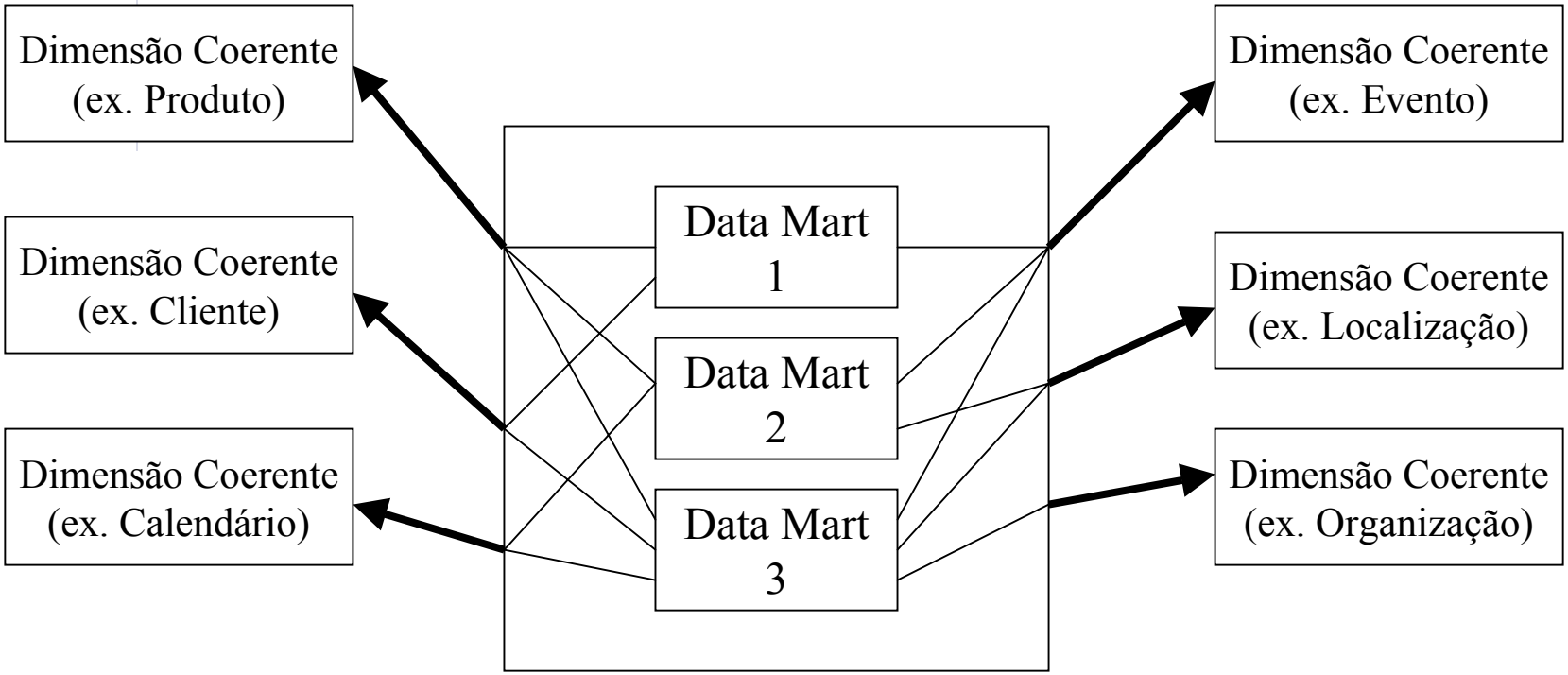
Modelação de Dados

- ✓ Resolução de Problemas de Incoerência de Dados
  - *Incompletude de Dados*
  - *Eventual incoerência de Dados - DW só pode ter uma versão da verdade!*
  - *Como resolver esta questão?*
    - Definições claras e horizontalmente partilhadas
    - Validação semântica dos conteúdos (Dados) contra as definições

DW = Oportunidade de definir "Metadata" (Dados sobre os Dados) Organizacional

Dimensões e Factos Coerentes

- ✓ Um DW com dimensões e factos coerentes são como um "bus interface" que permite a adição sucessiva de novos data marts



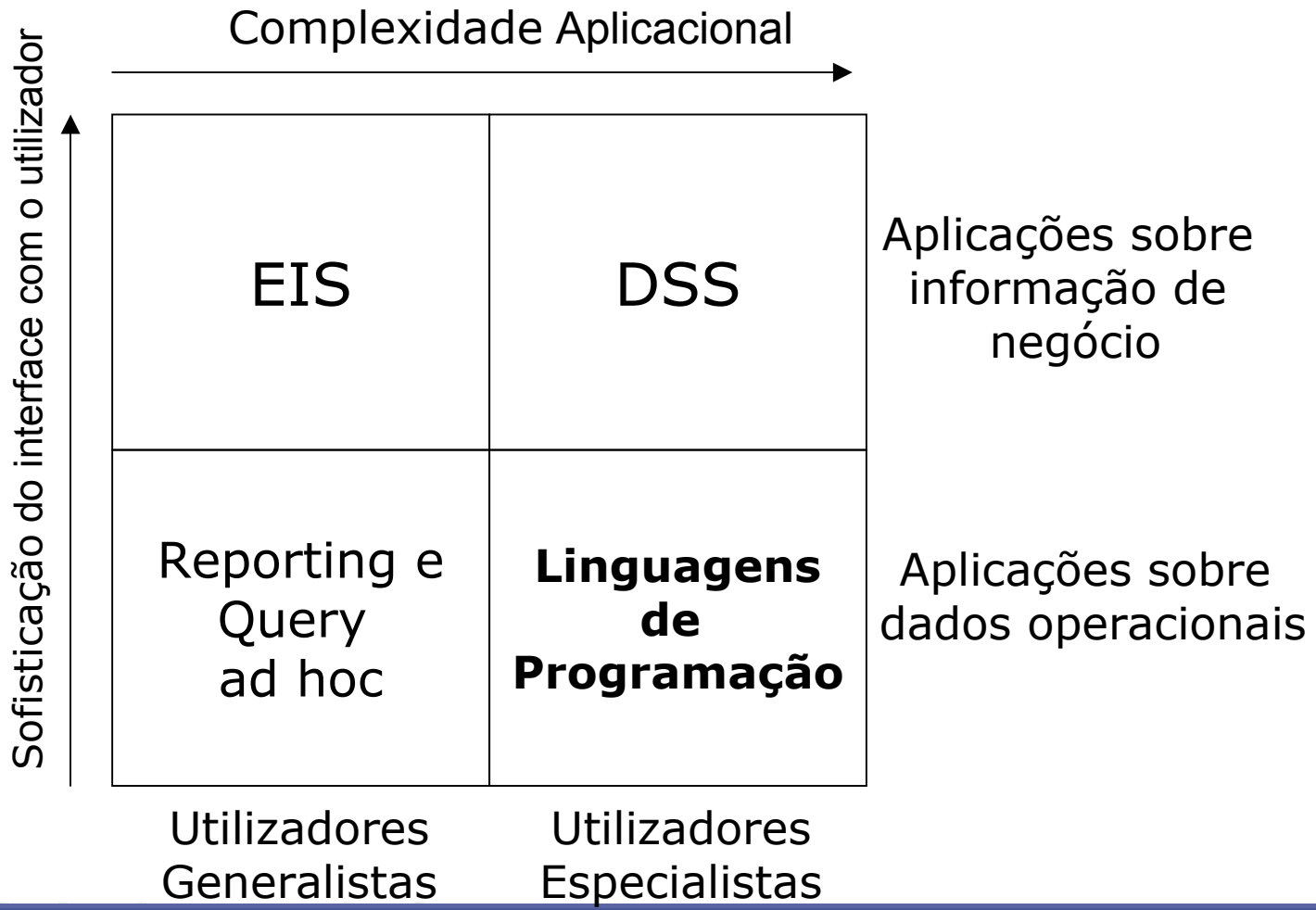
Dimensões Coerentes*

* "conformed dimensions"

- ✓ Dimensão coerente é uma dimensão que tem o mesmo significado qualquer que seja tabela de factos com a qual possa ser ligada
  - *Uma dimensão coerente é partilhada pelos diversos data marts que a referenciam*
  - *Exemplos: cliente, fornecedor, produto, tempo (calendário)*
  - *Do ponto de vista da consistência é uma das vantagens dos modelos ER aplicadas na modelagem multidimensional*
  - *Tornam possível a mesma interpretação do conceito e respectivos atributos ao longo dos diferentes data marts*
  - *Potenciam o cruzamento de informação de diferentes data marts*
  - *Representa 80% do esforço de modelagem*

Exploração de Dados

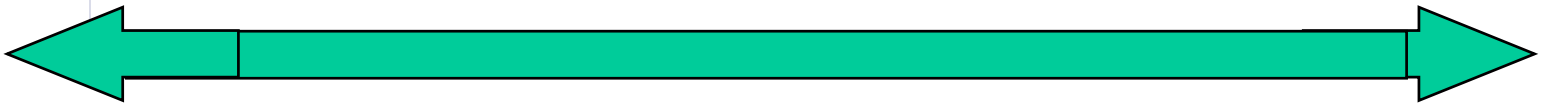
✓ Tipos de Ferramentas vs Tipos de Utilização



Exploração de Dados

- ✓ Papel do Analista de Negócio vs Processo de Análise dos Dados

Conduzido pelo Analista | **Assistido pelo Analista** | **Conduzido pelos Dados**



Processamento Informação
- *Query*
- *Reports*

Processamento Analítico
- OLAP Relacional
- OLAP Multidimensional

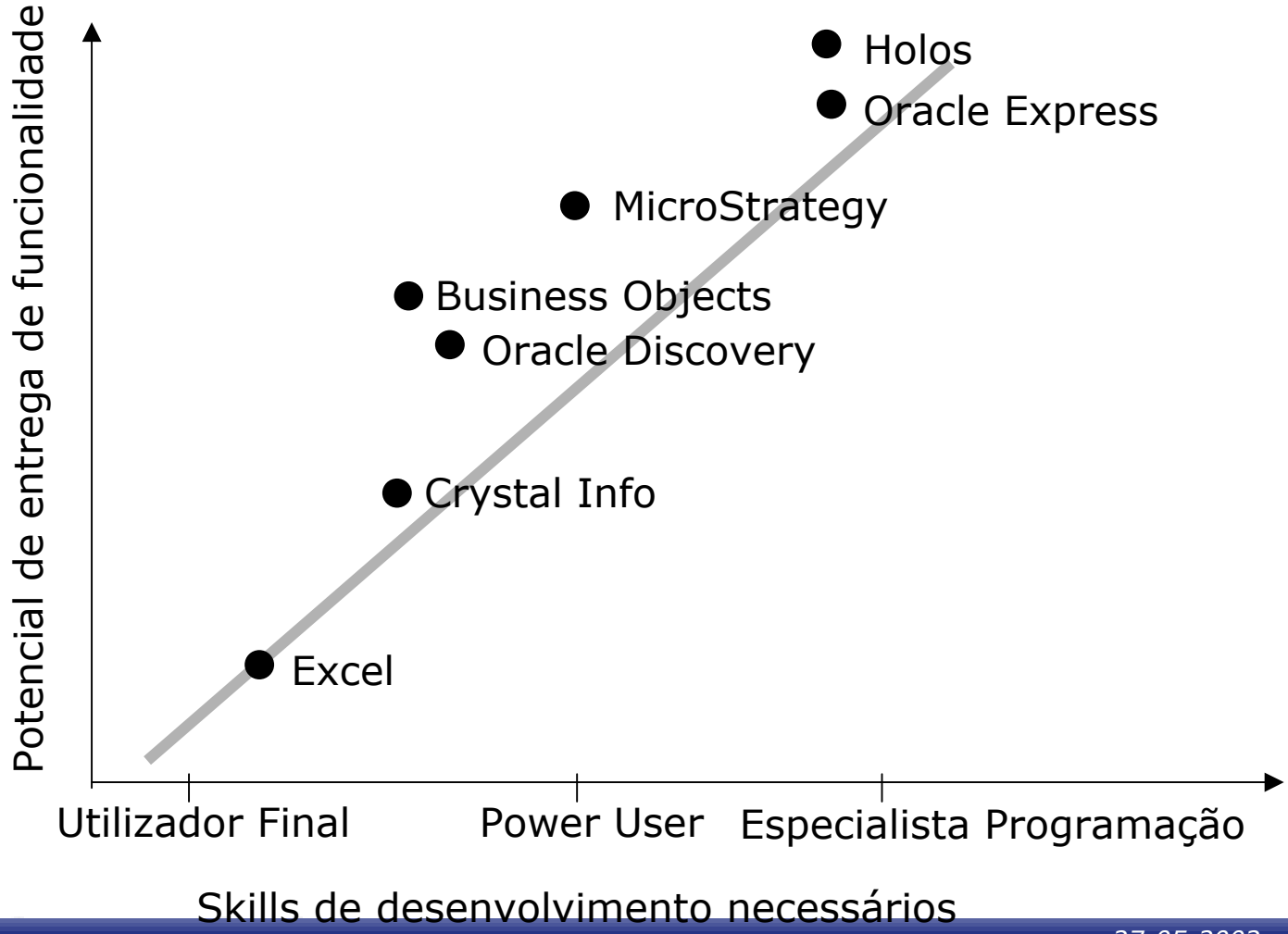
Métodos Estatísticos

Sistemas Periciais

Knowledge Discovery

Exploração de Dados

- ✓ Facilidade Desenvolvimento vs Potencial Funcionalidade



Exploração de Dados

- ✓ Algumas Características Importantes

  Interface

  - *Conjunto completo de funcionalidades disponíveis no interface com o utilizador (todas ou quase todas as disponíveis nas ferramentas de reporting e query ad hoc)*
  - *Capacidade de gerir alertas e gerar gráficos sofisticados*
  - *Sistema de navegação fácil e intuitivo. Capacidade do utilizador criar as suas próprias "bookmarks"*
  - *Suportar modelos de previsão*

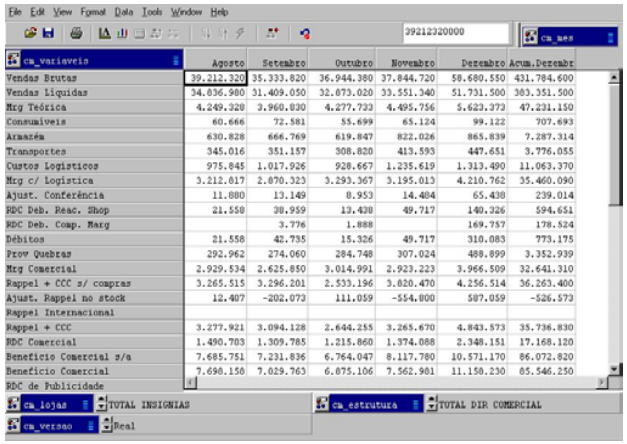
Exploração de Dados

✓ Base de Dados Multidimensional (MOLAP)

- *A monitorização de alto nível sobre o comportamento do negócio deve ser suportada por instrumentos eficientes*
- *A transformação dos dados operacionais em informação de negócio introduz forte sumarização (agregações de agregações)*
- *A representação multidimencional introduz vantagens de performance no acesso aos dados relativamente ao relacional.*
- *Algumas características importantes:*
  - 4GL potente para programação do carregamento e actualização dos cubos de dados
  - Metacubos ou cubos virtuais
  - Algoritmos adequados à dimensão e esparsidade dos cubos - cubos pequenos podem residir em memória principal

Exploração de Dados - Exemplos

- ✓ OLAP (MOLAP: Seagate Holos, Oracle Express, …)



cm_variaveis	Agosto	Setembro	Outubro	Novembro	Dezembro	Acum.Dezembr
Vendas Brutas	39.212.320	35.333.820	36.944.380	37.844.720	58.680.550	431.784.600
Vendas Liquidas	34.836.980	31.409.050	32.873.020	33.551.340	51.731.500	383.351.500
Mrg Teórica	4.249.320	3.960.030	4.277.733	4.495.756	5.623.373	47.231.150
Consumiveis	60.666	72.581	55.699	65.124	99.122	707.693
Armazém	630.828	666.769	619.847	822.026	865.839	7.287.314
Transportes	345.016	351.157	308.820	413.593	447.651	3.776.055
Custos Logisticos	975.845	1.017.926	928.667	1.235.619	1.313.490	11.063.370
Mrg c/ Logistica	3.212.617	2.870.323	3.293.367	3.195.013	4.210.762	35.460.090
Ajust. Conferência	11.880	13.149	8.953	14.484	65.438	239.014
RDC Deb. Reac. Shop	21.550	30.959	13.430	49.717	140.326	594.651
RDC Deb. Comp. Marg		3.776	1.888		169.757	178.524
Débitos	21.558	42.735	15.326	49.717	310.083	773.175
Prov Quebras	292.962	274.060	284.748	307.024	488.899	3.352.939
Mrg Comercial	2.929.534	2.625.850	3.014.991	2.923.223	3.966.509	32.641.310
Rappel + CCC s/ compras	3.265.515	3.296.201	2.533.196	3.020.470	4.256.514	36.263.400
Ajust. Rappel no stock	12.407	-202.073	111.059	-554.800	587.059	-526.573
Rappel Internacional						
Rappel + CCC	3.277.921	3.094.128	2.644.255	3.265.670	4.843.573	35.736.830
RDC Comercial	1.490.703	1.309.785	1.215.860	1.374.088	2.348.151	17.168.120
Beneficio Comercial s/a	7.685.751	7.231.836	6.764.047	8.117.780	10.571.170	86.072.820
Beneficio Comercial	7.698.158	7.029.763	6.875.106	7.562.901	11.158.230	85.546.250
RDC de Publicidade						

cm_lojas: TOTAL INSIGNIAS | cm_estrutura: TOTAL DIR COMERCIAL | cm_versao: Real

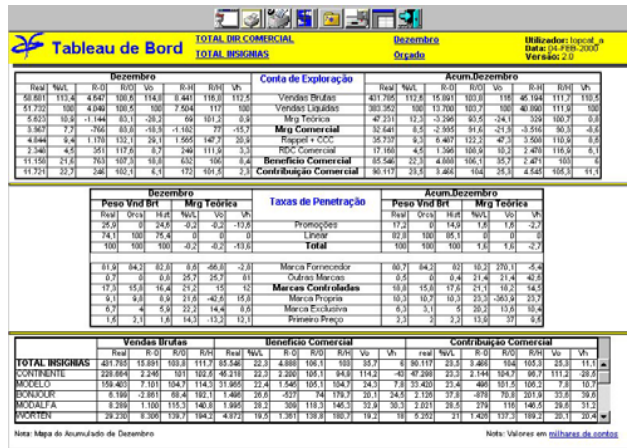


Tableau de Bord

TOTAL DIR COMERCIAL
TOTAL INSIGNIAS
Dezembro
Orçado

Utilizador: topcat_a
Data: 04-FEB-2000
Versão: 2.0

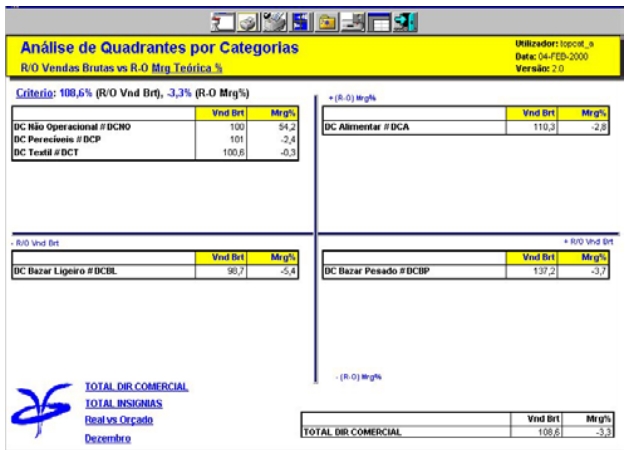
Conta de Exploração

Dezembro								Acum.Dezembro							
Real	%VL	R-O	R/O	Vo	R-H	R/H	Vh	Real	%VL	R-O	R/O	Vo	R-H	R/H	Vh
58.681	113,4	4.647	108,5	114,8	8.441	116,8	112,5	Vendas Brutas	431.785	112,6	15.891	103,8	116	46.194	111,7	110,5
51.732	100	4.046	100,5	100	7.504	117	100	Vendas Liquidas	383.352	100	13.700	100,7	100	40.990	111,9	100
5.623	10,9	-1.144	83,1	-28,3	60	101,2	0,9	Mrg Teórica	47.231	12,3	-3.298	83,5	-24,1	329	100,7	0,8
3.947	7,7	-766	81,8	-18,9	-1.182	77	-15,7	**Mrg Comercial**	32.641	8,5	-2.955	81,6	-21,5	-3.516	80,3	-8,6
4.844	9,4	1.176	132,1	29,1	1.565	147,3	20,9	Rappel + CCC	35.737	9,3	6.407	122,2	47,3	3.504	110,9	8,5
2.346	4,5	351	117,6	8,7	246	111,9	3,3	RDC Comercial	17.160	4,5	1.396	108,9	10,3	2.470	116,9	6,1
11.150	21,6	763	107,3	18,8	632	106	8,4	**Beneficio Comercial**	85.546	22,3	4.868	106,1	35,9	2.421	103	6
11.721	22,7	246	102,1	6,1	172	101,5	2,3	**Contribuição Comercial**	90.117	23,5	3.466	104	25,6	4.545	105,3	11,1

Dezembro							Acum.Dezembro						
Peso Vnd Brt		Mrg Teórica					Peso Vnd Brt		Mrg Teórica				
Real	Orçd	%VL	Vo	Vh			Real	Orçd	%VL	Vo	Vh		
25,9	0	24,6	-0,2	-0,2	-13,6	Promoções	17,2	0	14,9	1,6	1,6	-2,7	
74,1	100	75,4	0	0	0	Linear	82,8	100	85,1	0	0	0	
100	100	100	-0,2	-0,2	-13,6	**Total**	100	100	100	1,6	1,6	-2,7	
81,9	84,2	82,8	0,6	-66,8	-2,8	Marca Fornecedor	60,7	84,2	82	10,2	270,1	-5,4	
0,7	0	0,8	25,7	25,7	61	Outras Marcas	0,5	0	0,4	21,4	21,4	42,6	
17,3	15,8	16,4	21,2	-15	12	**Marcas Controladas**	16,8	15,8	17,6	21,1	10,2	14,5	
9,1	9,8	8,9	21,8	-42,8	15,8	Marca Propria	10,3	10,7	10,3	23,3	-363,9	23,7	
6,7	4	5,9	22,2	14,4	8,6	Marca Exclusiva	6,3	3,1	5	20,2	13,6	18,4	
1,6	2,1	1,6	14,3	-13,2	12,1	Primeiro Preço	2,3	2	2,2	13,6	37	9,6	

	Vendas Brutas				Beneficio Comercial				Contribuição Comercial							
	Real	R-O	R/O	R/H	Real	%VL	R-O	R/O	R/H	Vo	Vh	real	%VL	R-O	R/O	R/H	Vo	Vh
TOTAL INSIGNIAS	431.785	15.891	103,8	111,7	85.546	22,3	4.868	106,1	103	35,7	6	90.117	23,5	3.466	104	105,3	25,3	11,1
CONTINENTE	228.864	2.246	101	102,6	46.218	22,3	2.330	108,1	94,9	114,2	-40	47.298	23,3	2.144	104,7	99,7	111,2	-28,6
MODELO	159.403	7.101	104,7	114,3	31.965	22,4	1.546	105,1	104,7	34,3	7,8	33.400	23,4	490	101,5	106,3	7,8	10,7
BONJOUR	8.199	-2.861	68,4	102,1	1.466	26,6	-527	74	179,7	20,1	24,5	2.126	27,3	-879	70,8	201,9	33,8	39,6
MODALFA	8.289	1.100	115,3	140,8	1.995	28,2	309	118,3	146,3	32,9	30,3	2.021	28,5	279	116	146,5	29,8	31,2
WORTEN	29.230	8.306	139,7	194,2	4.872	19,5	1.361	139,8	180,7	19,2	18	5.262	21	1.408	137,3	189,2	20,1	20,4

Nota: Mapa do Acumulado de Dezembro
Nota: Valores em milhares de contos



Análise de Quadrantes por Categorias

R/O Vendas Brutas vs R-O Mrg Teórica %

Utilizador: topcat_a
Data: 04-FEB-2000
Versão: 2.0

Criterio: 108,6% (R/O Vnd Brt), -3,3% (R-O Mrg%)

	Vnd Brt	Mrg%
DC Não Operacional # DCNO	100	54,2
DC Perecíveis # DCP	101	-2,4
DC Textil # DCT	100,6	-0,3

+ (R-O) Mrg%

	Vnd Brt	Mrg%
DC Alimentar # DCA	110,3	-2,8

- R/O Vnd Brt

	Vnd Brt	Mrg%
DC Bazar Ligeiro # DCBL	98,7	-5,4

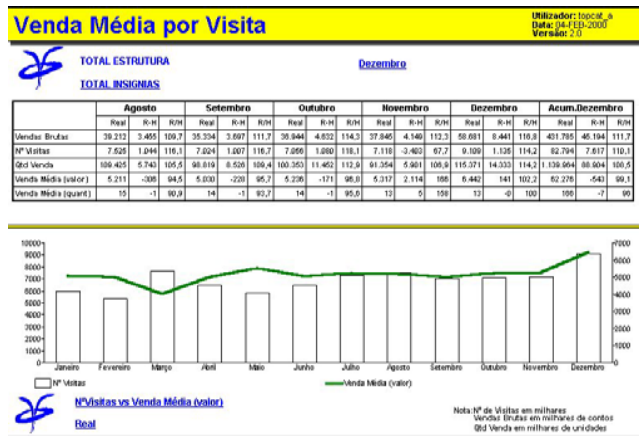
+ R/O Vnd Brt

	Vnd Brt	Mrg%
DC Bazar Pesado # DCBP	137,2	-3,7

- (R-O) Mrg%

	Vnd Brt	Mrg%
TOTAL DIR COMERCIAL	108,6	-3,3

TOTAL DIR COMERCIAL
TOTAL INSIGNIAS
Real vs Orçado
Dezembro



Venda Média por Visita

Utilizador: topcat_a
Data: 04-FEB-2000
Versão: 2.0

TOTAL ESTRUTURA
TOTAL INSIGNIAS
Dezembro

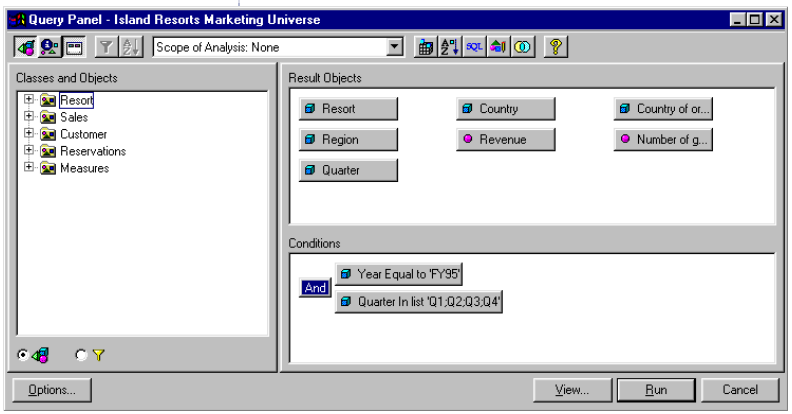
	Agosto			Setembro			Outubro			Novembro			Dezembro			Acum.Dezembro		
	Real	R-H	R/H	Real	R-H	R/H	Real	R-H	R/H	Real	R-H	R/H	Real	R-H	R/H	Real	R-H	R/H
Vendas Brutas	39.212	3.455	109,7	35.334	3.697	111,7	36.944	4.832	114,3	37.845	4.149	112,3	58.681	8.441	116,8	431.785	46.194	111,7
Nº Visitas	7.525	1.044	116,1	7.024	1.007	116,7	7.056	1.060	118,1	7.118	-3.483	67,7	9.109	1.126	114,2	82.794	7.617	110,1
Qtd Venda	109.425	5.743	105,5	98.619	8.526	109,4	100.353	11.462	112,9	91.354	5.961	106,9	115.371	14.333	114,3	1.139.664	88.904	100,5
Venda Média (valor)	5.211	-308	94,5	5.030	-220	95,7	5.236	-171	96,8	5.317	2.114	166	6.442	141	102,2	82.276	-543	99,3
Venda Média (quant)	15	-1	90,9	14	-1	93,7	14	-1	95,6	13	5	159	13	-0	100	166	-7	99

Nº Visitas vs Venda Média (valor)
Real

Nota: Nº de Visitas em milhares
Vendas Brutas em milhares de contos
Qtd Venda em milhares de unidades

Exploração de Dados - Exemplos

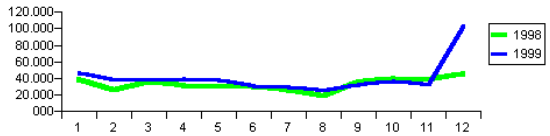
- ✓ Ad-Hoc Query (MicroStrategy, Brio, Business Objects, Oracle Discovery, …)



Análise Evolutiva de Vendas

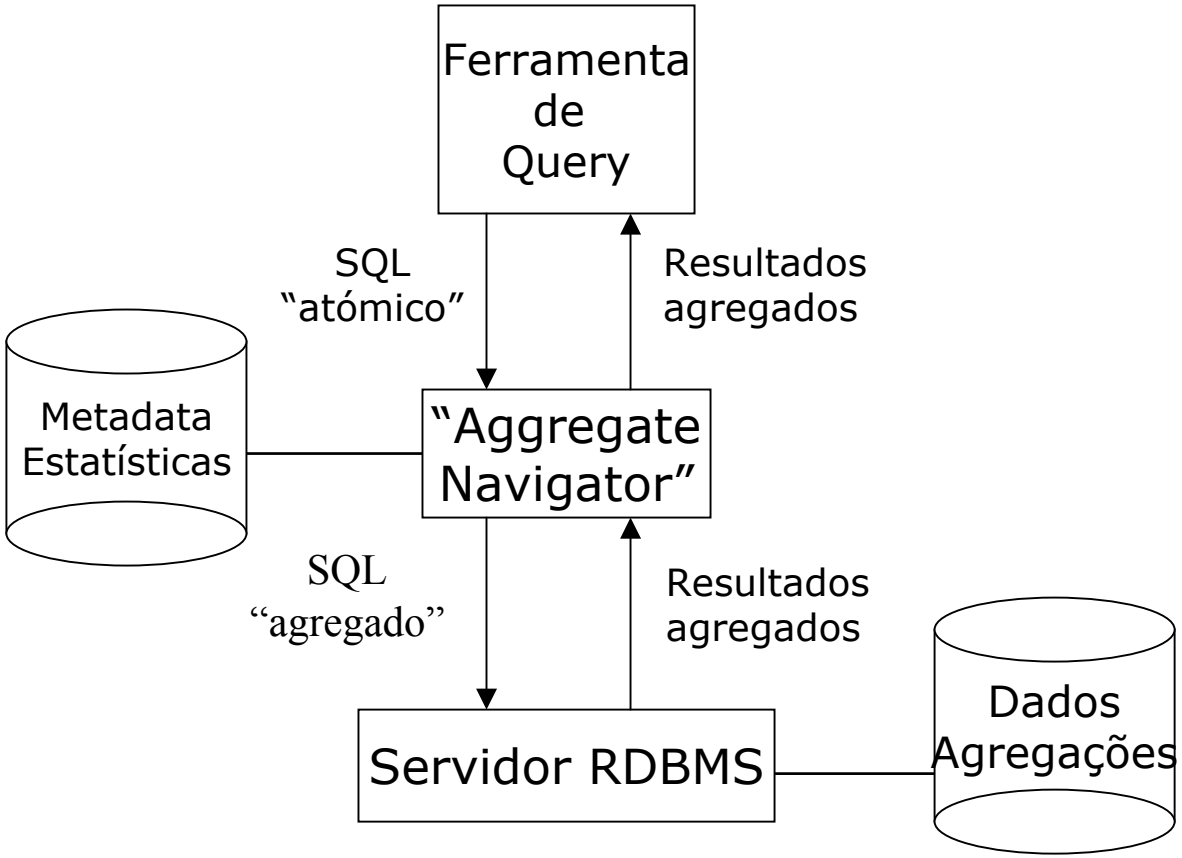
BASE MOLH.MAIZENA EXPRES.250G CX (002.000.022)

Ano	Mês	VND BRT
1998	1	39.004
1998	2	24.864
1998	3	35.616
1998	4	31.584
1998	5	30.240
1998	6	29.904
1998	7	25.200
1998	8	17.808
1998	9	35.280
1998	10	39.984
1998	11	37.968
1998	12	45.696
1999	1	47.376
1999	2	37.632
1999	3	38.976
1999	4	40.320
1999	5	38.304
1999	6	31.248
1999	7	29.376



Explorações de Dados

- ✓ ROLAP => "Aggregate awareness"



Sumário: Razões para construir um DW

- ✓ Convergência / Visibilidade num ponto único
- ✓ Fácil navegabilidade
- ✓ Separação Operacional / Analítica
  - *Não impacto de queries nas operações*
  - *Modelação diferente*
- ✓ Resolução de problemas de coerência de definições e de dados
- ✓ Sustentação de novas aplicações de Business Intelligence

Processos base de um DW

- ✓ Extracção de Dados
- ✓ Transformação:

  Limpeza, Reformatação, Combinação
- ✓ Carregamento
- ✓ Controlo de Qualidade
- ✓ Publicação; Actualização/Refresh; Interrogação
- ✓ Auditoria
- ✓ Back-Up e Recuperação

A Função de DW Manager

- ✓ Definição da Função
  - *"Guardião do Templo"*
  - *"Editor" responsável pela qualidade dos dados publicados*
  - *Responsável pela Metadata Organizacional*
  - *Facilitador da prioritização do desenvolvimento de todas as aplicações de Suporte à Decisão*
  - *Responsável pela Publicação, nos timings acordados, das novas versões dos dados*

A Função de DW Manager

- ✓ Valências exigidas pela Função
  - *Conhecimento do Negócio (em particular, das necessidades dos "Knowledge Workers")*
  - *Capacidade de Comunicação e Facilitação / Geração de Consenso*
  - *Capacidade Organizativa e Disciplina de Entrega (c/Controlo de Qualidade)*
  - *Conhecimentos Técnicos específicos de DW / DSS / BI*
  - *Resiliência (DW é um Processo!)*
  - *Capacidade de Modelagem Avançada*

Dimensão de um DW

Qual o tamanho da "Piscina" ?

Dados

- **Números de Tabelas**
- **Volume de Dados**
- **Número de Fontes operacionais (instâncias)**
- **Número de Data-Marts (instâncias)**

Processos

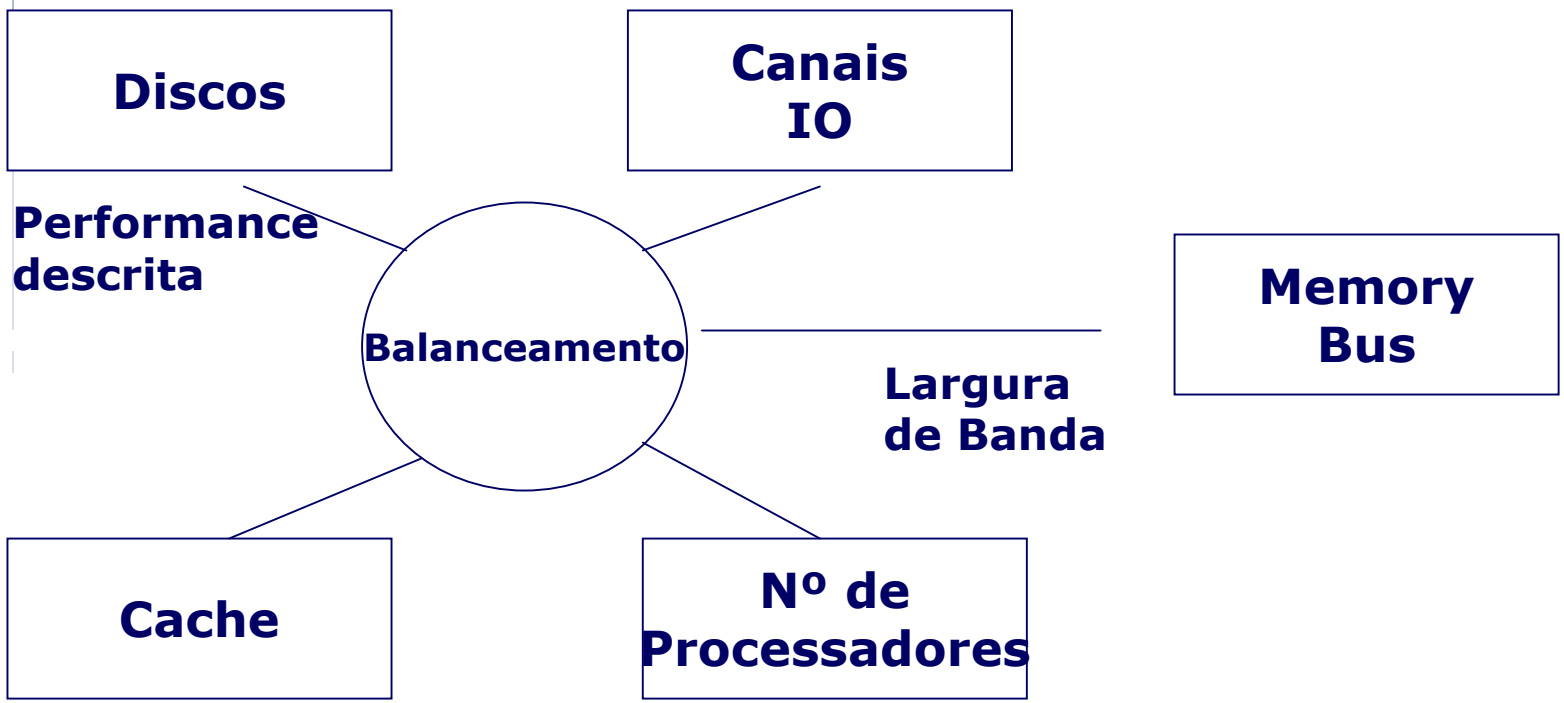
- **Número de Processos distintos do Batch**
- **Número de Queries por dia (atendidos / não atendidos)**

IMPACTO NA INFRAESTRUTURA DE SERVIDORES!

Problemas Técnicos que se levantam

- ✓ Distribuição da computação SMP ou MPP?
- ✓ Como diminuir ao máximo a contenção entre processos?
- ✓ Como facilitar a paralelização implícita do software?
- ✓ Como garantir Back-Up's íntegros e prontos a suportar uma recuperação?
- ✓ Como implementar "Disaster Recovery"?
- ✓ Como garantir que as operações pouco extensas de delete/update não têm grande impacto de performance?
- ✓ Como promover o balanceamento dinâmico de carga entre processadores?

Balanceamento da Arquitectura Tecnológica como um todo



Pensar Arquitectura Tecnológica DW … a tempo

A escalabilidade não acontece: Arquitecta-se!

- ✓ DW: A explosão de utilização é sempre mais rápida do que o previsto
- ✓ Atenção à "Procura Escondida" (reprimida)
- ✓ Necessidade de revisão da Arquitectura Tecnológica é normal

Leituras recomendadas

- ✓ "The Data Warehouse, Lifecycle Toolkit", Ralph Kimball