

TECNOLOGIA E DESENVOLVIMENTO DE SISTEMAS

FUNDAMENTOS: BUSINESS INTELLIGENCE E DATA WAREHOUSE

PROF. JORGE SURIAN jorge.surian@gmail.com
PROF. RENATO KIMURA kimura.renato@gmail.com
PROF. SALVIO PADLIPSKAS salvio@fiap.com.br

CARACTERÍSTICAS DO AMBIENTE OLTP E OLAP

OLTP: *ON-LINE TRANSACTION PROCESSING*

- Processamentos que executam as operações do dia-a-dia da organização.
- Ênfase ao suporte do negócio, através de um processamento rápido, acurado e eficiente dos dados.
- Exemplos: Controle de Pedidos e Faturamento; Administração de Pessoal.

OLAP: *ON-LINE ANALYTICAL PROCESSING*

- Processamentos que suportam a tomada de decisões.
- Permite analisar tendências e padrões em grandes quantidades de dados ao longo do tempo (histórico) e em diferentes localizações (geográficos).

	<i>BD Operacional</i>	<i>Data Warehouse</i>
<i>Usuários</i>	Funcionários	Alta Administração
<i>Utilização</i>	Tarefas Cotidianas	Decisões Estratégicas
<i>Padrão de Uso</i>	Previsível	Difícil de Prever
<i>Funcionamento</i>	Baseado em Transações	Baseado em Análise de Dados
<i>Valores dos Dados</i>	Valores Atuais. Voláteis	Valores Históricos. Imutáveis
<i>Detalhamento</i>	Alto	Sumarizado
<i>Organização dos Dados</i>	Orientado à Aplicação	Orientado ao Assunto

AMBIENTE DE SUPORTE A DECISÃO

DSS (*Decision Support System*)

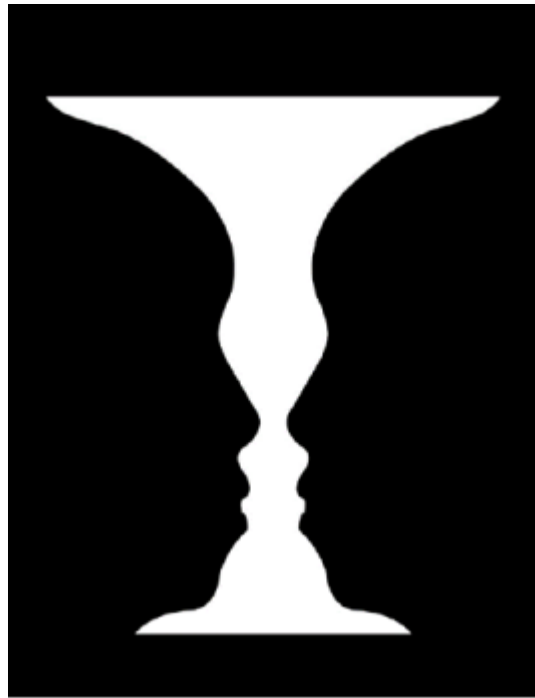
- O objetivo é prover um ambiente que armazene, integre, disponibilize e compartilhe informações relevantes, com qualidade, agilidade e flexibilidade visando reagir as tomadas de decisão.
- Facilita a identificação de ameaças e oportunidades de negócios, para que oportunidades sejam aproveitadas em primeira mão, ameaças sejam convertidas em boas chances de lucro e a capacidade de resposta à mudança seja ampliada, junto com o dinamismo característico do mercado.

AMBIENTE DE SUPORTE A DECISÃO

Atualmente temos:

- Maior concorrência
- Fusões / Aquisições
- Parcerias estratégicas

Se sobressai quem consegue enxergar além do básico...



BUSINESS INTELLIGENCE

Conceitos



- Descreve as habilidades das empresas em explorar informações, desenvolvendo percepções e entendimento a seu respeito, proporcionando uma melhor tomada de decisão.
- As ferramentas de BI trabalham em conjunto com o *Data Warehouse (DW)*, fornecendo informações para a tomada de decisão.
- O *Data Warehouse* pode ser considerado um componente de BI.



ALGUNS OBJETIVOS DO BI

- Facilitar o acesso das informações aos usuários.
- Apoiar a organização com dados para a tomada de decisões.
- Proporcionar análises visando entender a necessidade dos clientes e da empresa.
- Definir e acompanhar estratégias para a organização.

TIPOS DE BI

- **ESTRATÉGICO**

Auxilia executivos e a área estratégica a desenvolver e apurar o progresso dos objetivos de longo prazo (aumento de vendas, lucros, redução de custos, retenção de clientes, etc.).

- **TÁTICO**

Análise de esforços de curto espaço de tempo focado em departamentos (marketing, vendas, serviços, finanças, etc.). Os objetivos são específicos e, normalmente, relacionados a linha de negócios (LOB).

- **OPERACIONAL**

Gerenciar o dia a dia dos negócios e processos. Definição, gerenciando o desempenho pelas operações diárias da organização. Tem como objetivo reduzir o tempo entre o fato e a tomada de decisão.

PILARES DO *DATA WAREHOUSE*

- Termo concebido por Bill Inmon, considerado o pai desta tecnologia.
- “Um *Data Warehouse* é uma coleção de dados, **orientado a assuntos, integrado, variável em relação ao tempo** e **não volátil**, para suportar o gerenciamento dos processos de tomada de decisão.” *Bill Inmon*
- Também conhecido como Armazém de Dados, projetado especialmente para organizar os dados de tal forma que facilite e viabilize o acesso a informações, o que não é possível no modelo tradicional de armazenamento de dados.

DW: ORIENTADO AO ASSUNTO

ORIENTAÇÃO AO ASSUNTO

Os dados em um DW são organizados de maneira a facilitar sua análise, ou seja, se prioriza os aspectos interpretativos e não os operacionais.

Os dados são organizados por assuntos e não por aplicações, como ocorre nos bancos de dados operacionais.



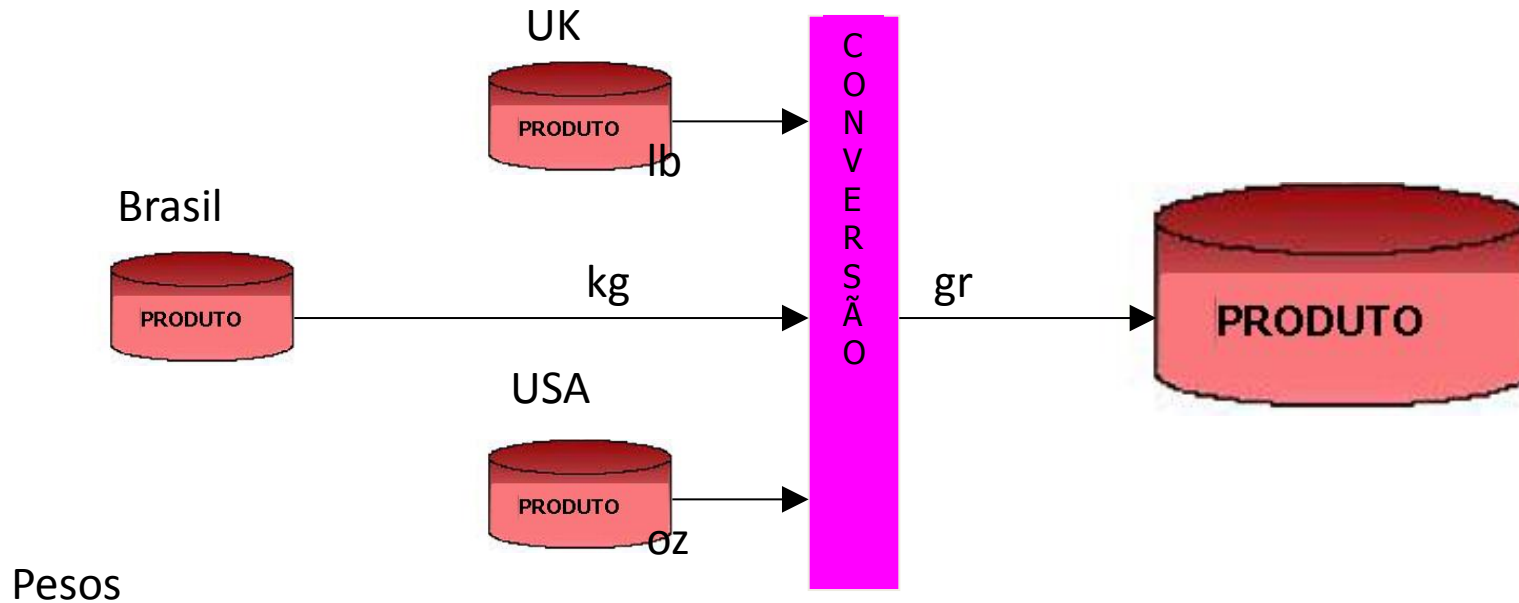
DW: INTEGRADO

INTEGRAÇÃO DE DADOS

Dados de um DW tem distintas origens.

Devem ser sumarizados ou eliminados

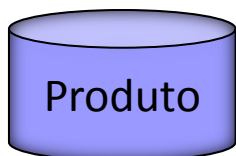
Devem ser padronizados (nomes, unidades de medidas, sexo)



DW: VARIÁVEL EM RELAÇÃO AO TEMPO

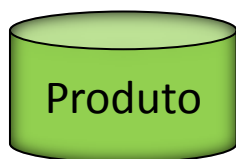
Enquanto que no BD o dado se refere a data de hoje, no DW o dado varia em relação ao tempo.

Transacional



Produto	Preço Vigente
Borracha	R\$ 3,50
Caderno	R\$ 12,00
Caneta	R\$ 2,50
Lápis	R\$ 1,70

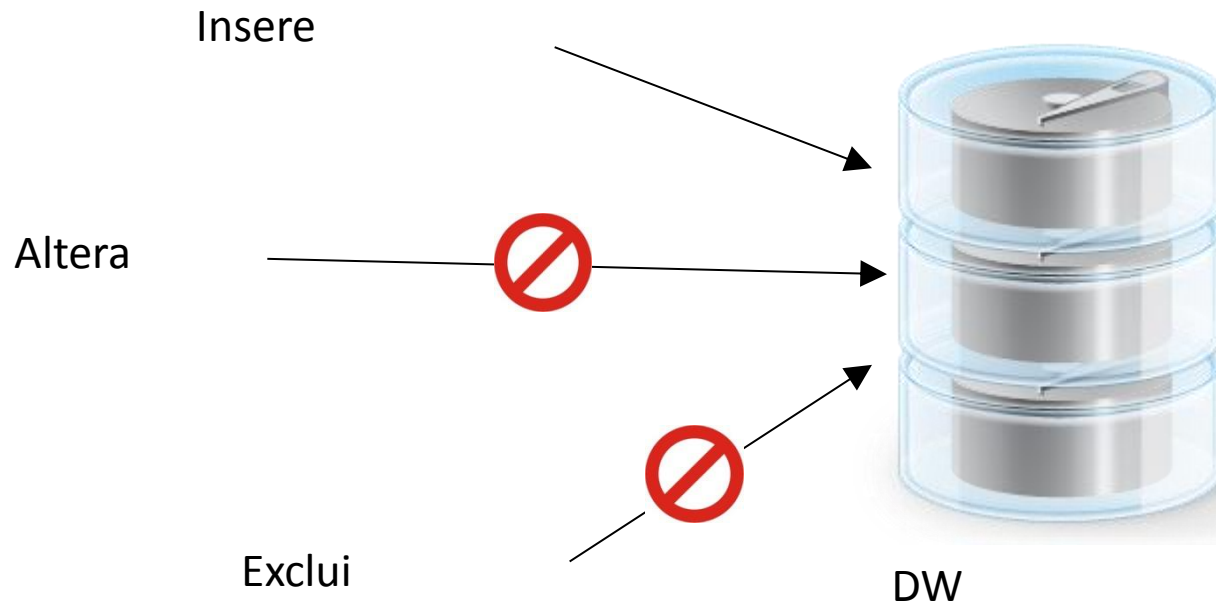
DW



Produto	Janeiro 2012	Março 2013	Janeiro 2014	Fevereiro 2015
Borracha	R\$ 2,50	R\$ 2,80	R\$ 3,10	R\$ 3,50
Caderno	R\$ 8,50	R\$ 9,40	R\$ 11,20	R\$ 12,00
Caneta	R\$ 1,50	R\$ 1,90	R\$ 2,10	R\$ 2,50
Lápis	R\$ 0,70	R\$ 0,95	R\$ 1,50	R\$ 1,70

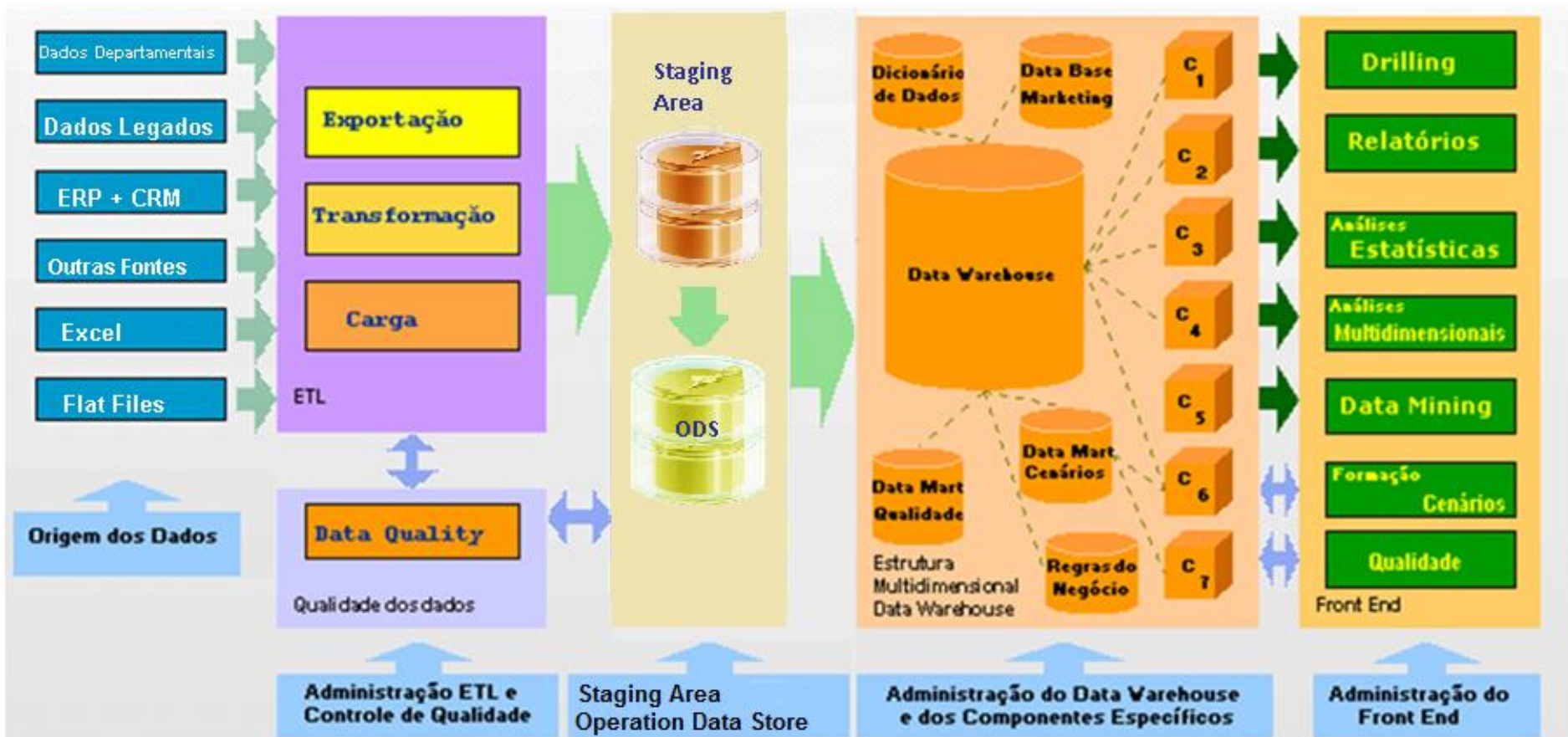
DW: NÃO VOLÁTIL

- Nenhum dado pode ser alterado depois de incluído no DW.
- Em um BD operacional podemos inserir, alterar e excluir um ou mais registros.
- Em um DW somente inserções são possíveis.
- Garante-se que em um DW, qualquer consulta a um dado relativo a um período de tempo, sempre produza o mesmo resultado



ARQUITETURA TÍPICA DE UM BI

ADAPTADA AS IDEIAS DE INMON



ARQUITETURA TÍPICA DE UM BI

Origem dos dados

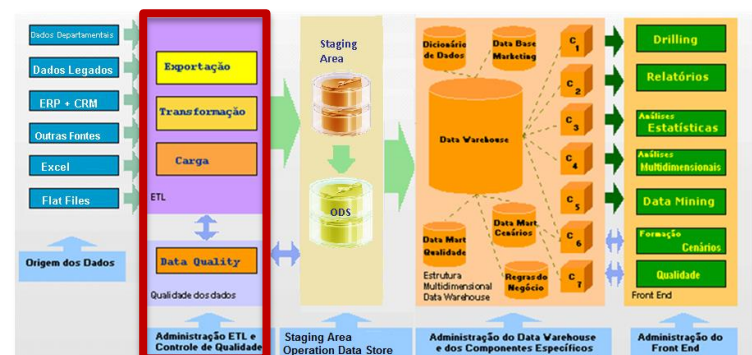
- Inúmeras fontes dos mais diversos tipos
- Bancos de dados operacionais (satélites, ERP)
- Fontes externas:
 - Flat Files com informações relevantes a necessidade do negócio
 - Geralmente desestruturados.



ARQUITETURA TÍPICA DE UM BI

ETL

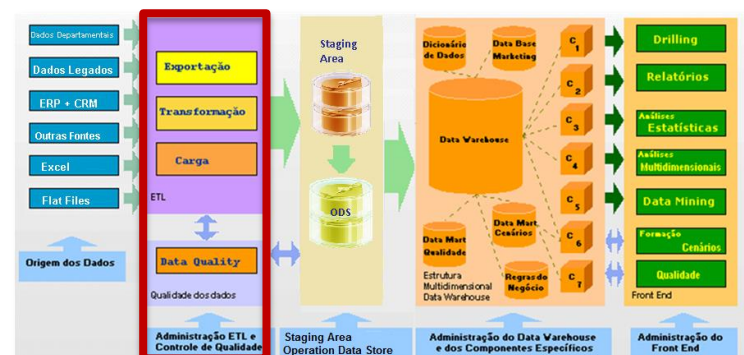
- *Extract, Transform & Load* é considerado por muitos um processo indispensável na construção de um DW.
- Tem como proposta extrair os dados na origem, transformar e carregar para o *Data Warehouse* e *Data Mart*.
- Essa etapa ocorre após a Arquitetura de DW / BI estar definida e implementada, integrando-se totalmente com a área de *Staging* e DW.
- A proposta é realizar a Extração dos dados puros, ou seja, como estão na origem, incluindo dados incorretos. Em resumo aqui ocorre leitura na origem e gravação na staging (Insert e Update)



ARQUITETURA TÍPICA DE UM BI

ETL

- Fase extremamente critica de um Data Warehouse, pois envolve a movimentação dos dados de origem nos sistemas transacionais e/ou sistemas legados, obedecendo as regras de negócio.
- O primeiro passo a ser tomado no processo de ETL: a definição das fontes de dados e fazer a extração deles.
- A extração de dados a partir das origens de dados para o ambiente de DW geralmente demanda mudanças de tecnologia (BD hierárquicos, Mainframes, entre outros).
- Um importante fator que deve ser levado em consideração é que dificilmente há o modelo de dados dos sistemas antigos, e se existem não estão atualizados e devidamente documentados.



ARQUITETURA TÍPICA DE UM BI

Staging Area

- É uma área de trabalho intermediária com o objetivo de centralizar os dados dos ambientes transacionais (OLTP) e das diferentes fontes existentes.
- A proposta é integrar todas as fontes de dados em um único ponto evitando o conceito de “ilhas” de dados.
- Como o volume de dados geralmente é muito grande, há muitos casos que não temos condições de processar as extrações e transformações na janela de tempo em que o DW não está sendo usado, sendo necessário fazer uso do que chamamos de staging área para conseguirmos executar os processos com sucesso.
- A área de staging é volátil, pois os dados armazenados variam ao longo do tempo e sofrem alterações incrementais.



ARQUITETURA TÍPICA DE UM BI

Staging Area

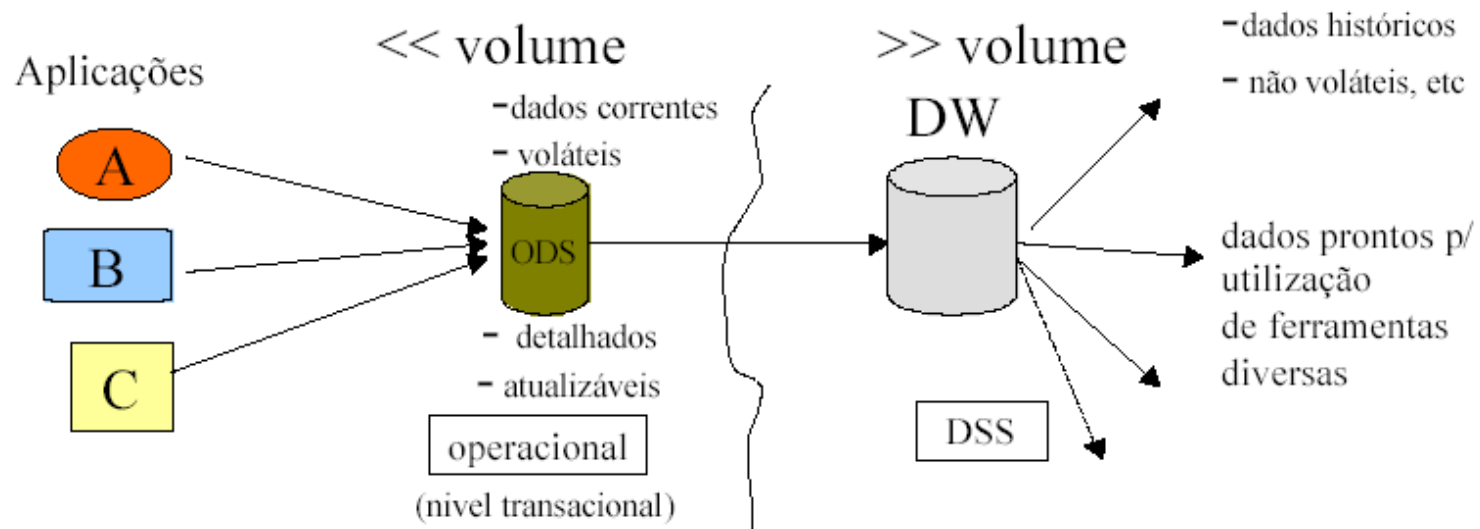
- A Staging Area é considerada área fora do acesso dos usuários.
- A Staging Area não deve suportar consultas dos usuários.
- Não tem como objetivo ser utilizada para consulta, mas sim como área de limpeza e transformações para as demais áreas do DW.
- A área de *Staging*, por geralmente ter granularidade idêntica aos sistemas operacionais, também serve para análises desse ambiente. Nessa situação não ocorre nenhuma sumarização, mas apenas conversões e consolidações que irão proporcionar melhor desempenho na fase de atualização dos dados.
- O Modelo de dados não precisa ser modelado segundo uma técnica específica. Os dados são armazenados muito próximo ao seu formato original.



ARQUITETURA TÍPICA DE UM BI

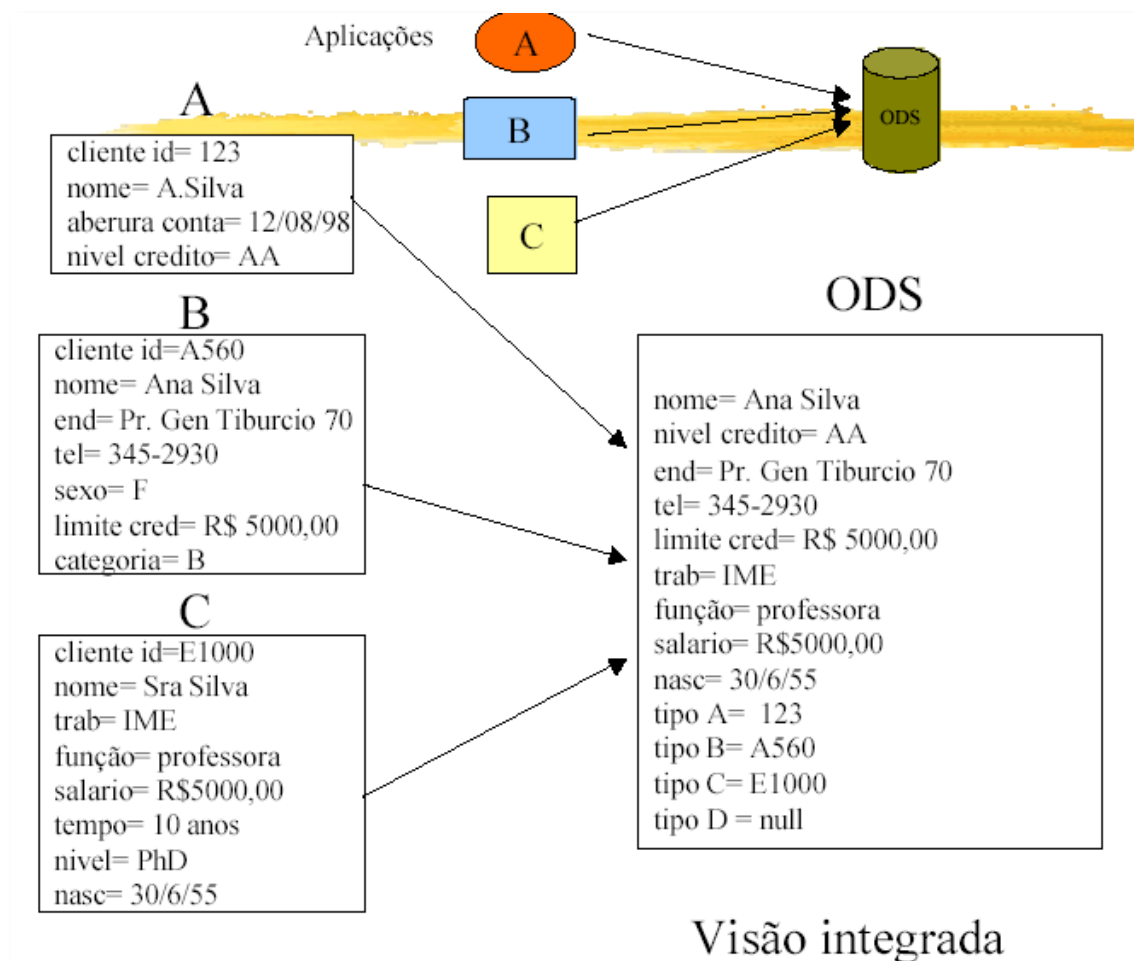
Staging Area: ODS (Operational Data Store)

- Armazenamento de dados atuais ou quase atuais para suporte à decisões mais iminentes: saldos, novos pedidos, fabricação de novos produtos entre outros.
- É orientado a assunto, é integrado, porém é volátil, ou seja, permite atualizações
- Mostra a visão corporativa dos dados em nível detalhado.
- É utilizado quando há a necessidade de se integrar os dados no nível operacional
- Dados mais antigos podem ser movidos e sumarizados para o DW



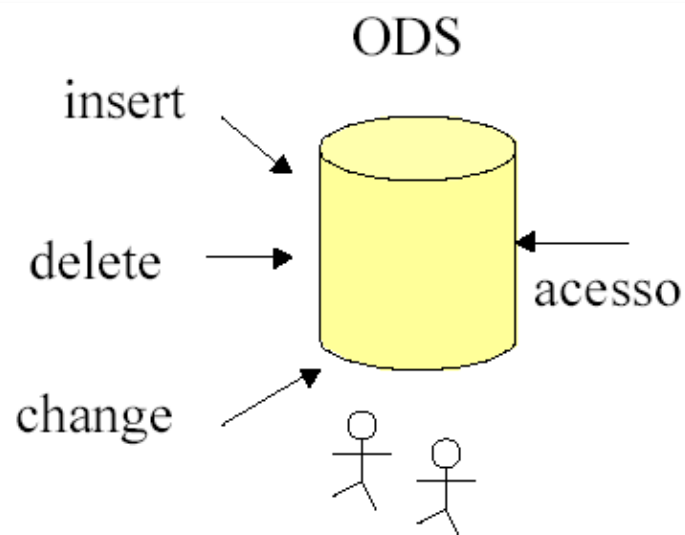
ARQUITETURA TÍPICA DE UM BI

Staging Area: ODS (Operational Data Store)

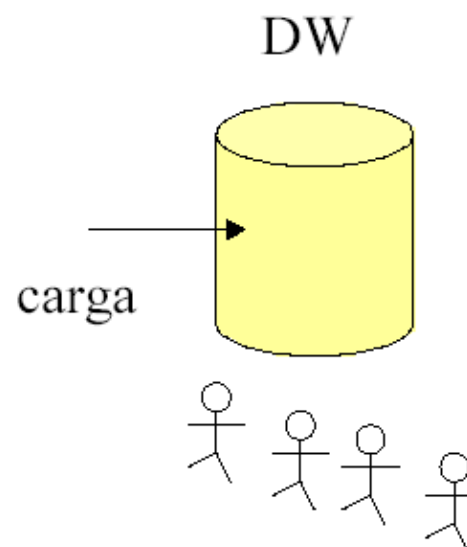


ARQUITETURA TÍPICA DE UM BI

Staging Area: ODS (Operational Data Store)



- propósito geral
- tempo de resposta rápido
- dia a dia /decisões pequeno grupo



- carga/acesso
- nenhuma atualização
- decisões a longo termo
- audiência gerencial

ARQUITETURA TÍPICA DE UM BI

Administração do *Data Warehouse*

- A camada *Data Warehouse* reúne, em uma **estrutura estrela** ou **flocos de neve**, os dados que foram extraídos, transformados e carregados das bases operacionais.
- A camada dos *Data Marts* permite acesso descentralizado aos dados.
- Os *Data Marts* são subconjuntos do *Data Warehouse* e geralmente fornecem informações direcionadas a um assunto específico.



ARQUITETURA TÍPICA DE UM BI

Administração do *Data Warehouse*

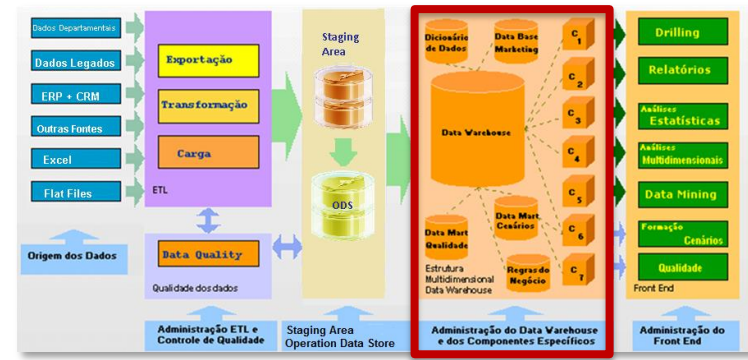
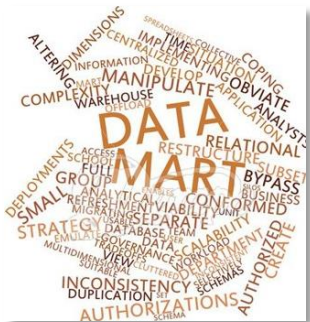
- Um *Data Warehouse* não é um “produto pronto” para ser instalado nas corporações, ou seja, precisa ser construído especificamente para cada organização.
- Sua construção requer profundo conhecimento das informações geradas do ambiente organizacional, sendo extremamente recomendado o apoio executivo.



DATA MART

Conceitos

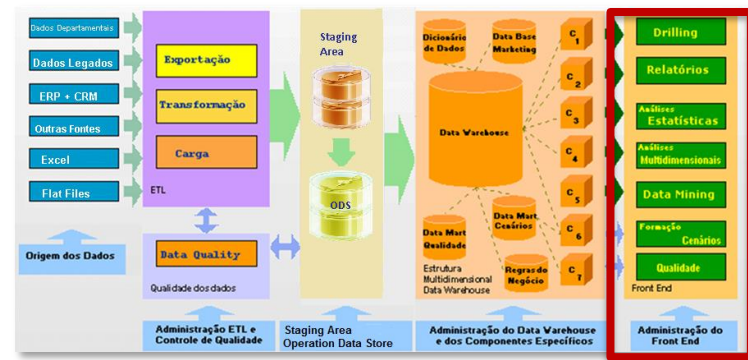
- Geralmente está relacionado a um sub conjunto de dados de um DW, referenciando-se a um assunto em especial.
- Em termos conceituais, pode-se afirmar que um *Data Mart* é um pequeno *Data Warehouse* que fornece suporte a decisão para um pequeno grupo de pessoas.
- Os dados contidos em um *Data Mart* podem ser diretamente carregados do DW ou da área de *Staging*, de acordo com a estratégia estabelecida.



ARQUITETURA TÍPICA DE UM BI

OLAP (Online Analytical Processing)

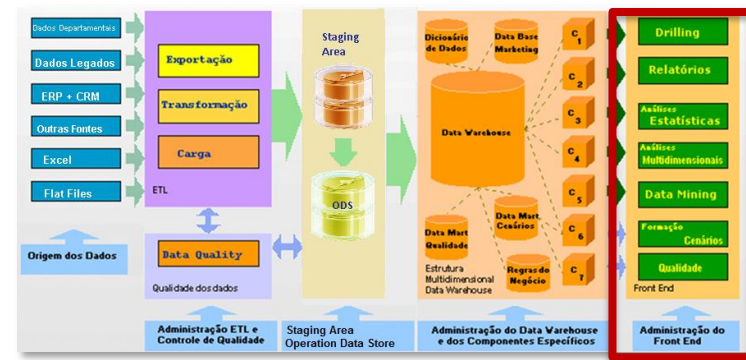
- Monitoramento dos fatos e indicadores mais relevantes da organização sintetizados em painéis de controle e relatórios executivos.
- Investigações *ad-hoc* e estudos estatísticos e matemáticos sobre os fatos da organização.
- Natureza das consultas:
 - Agregação (total de vendas, porcentagem em relação ao total)
 - Comparação Orçado vs Realizado
 - Ranking Top 10
 - Análise de quartil ou outros períodos de tempo estabelecidos
 - Visualização



ARQUITETURA TÍPICA DE UM BI

OLAP (Online Analytical Processing)

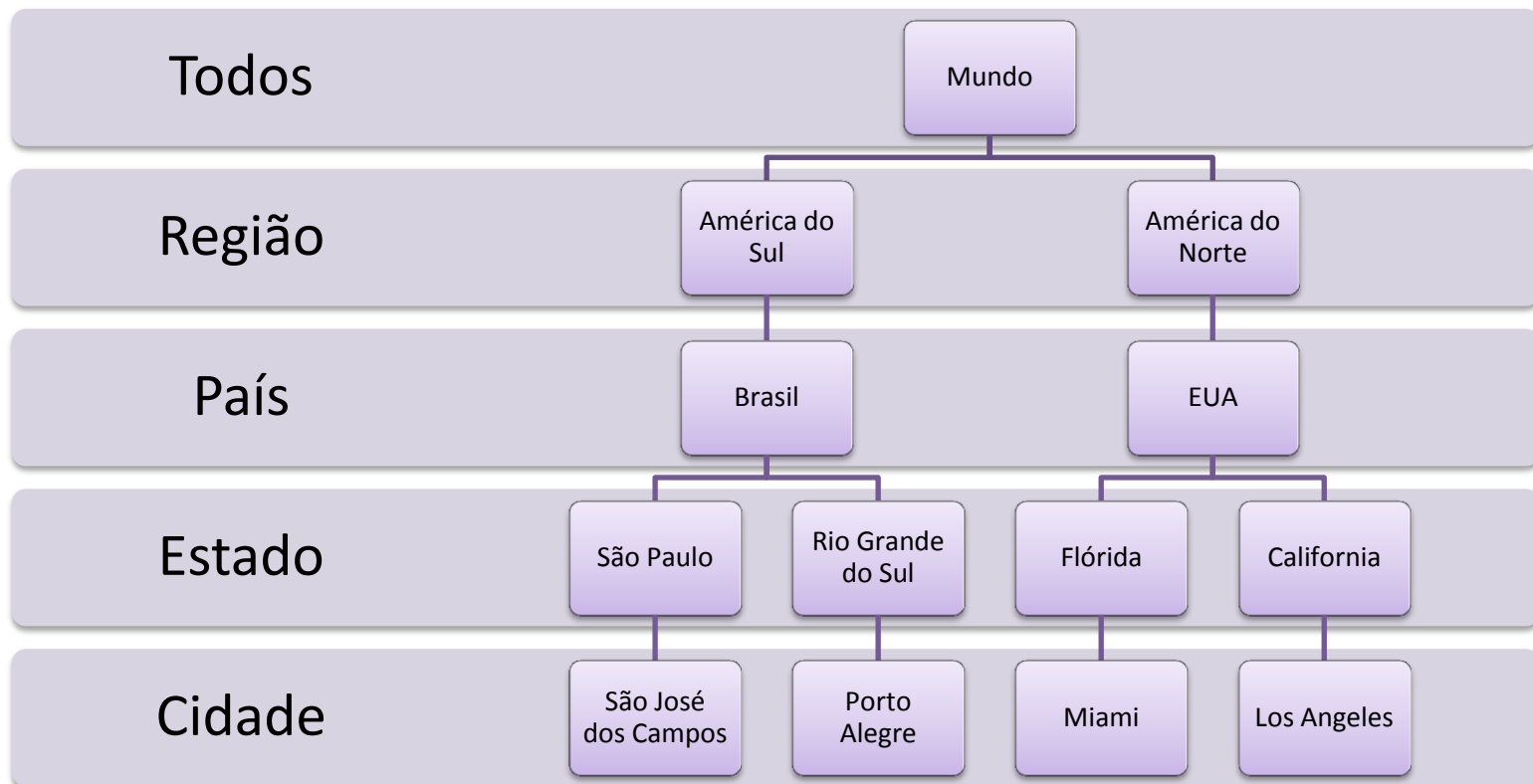
- Termos específicos utilizados como *pivot tables*, *drill-down* e *roll-up*, *slice* e *dice* entre outros.



ARQUITETURA TÍPICA DE UM BI

O que são Hierarquias e Níveis?

- É um conjunto de atributos que podem ser combinados de forma que possa fazer um drill down ou drill up na informação. Já o nível, é um atributo dentro de uma hierarquia



ARQUITETURA TÍPICA DE UM BI

Drilling: Conceito de Agrupamento

Eletronico.cdd - IBM Cognos Insight

Estilo Inserir Obter Dados Explorar

NmLinha Geral Measures Total de NmCliente Total de UF Total de NmVendedor

	Qtde	Preco	PrecoCMS
Total de NmLinha	8.168	R\$ 7.187.324	R\$ 8.866.237
+ ELETRODOMESTICOS	2.329	R\$ 2.311.118	R\$ 2.819.590
+ TI	3.166	R\$ 2.605.301	R\$ 3.246.926
+ SOM & IMAGEM	1.378	R\$ 1.145.022	R\$ 1.416.896
+ TELEFONIA	1.295	R\$ 1.125.882	R\$ 1.382.825

ELETRODOMESTICOS	2.329	R\$ 2.311.118	R\$ 2.819.590
+ ELETRODOMESTICOS, LINHA BRANCA	509	R\$ 367.588	R\$ 456.072
+ ELETRODOMESTICOS, ELETRODOMESTICOS DE COZINHA	1.076	R\$ 1.353.346	R\$ 1.639.292
+ ELETRODOMESTICOS, APARELHOS DE AR	744	R\$ 590.184	R\$ 724.227

ARQUITETURA TÍPICA DE UM BI

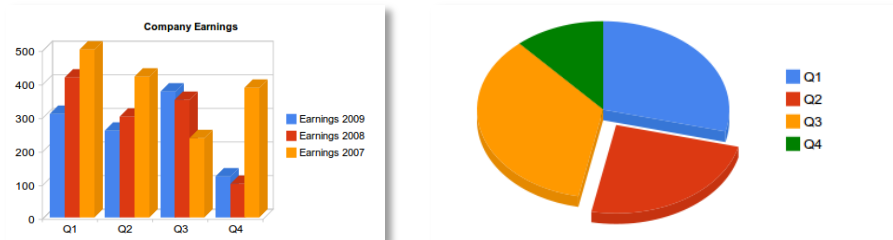
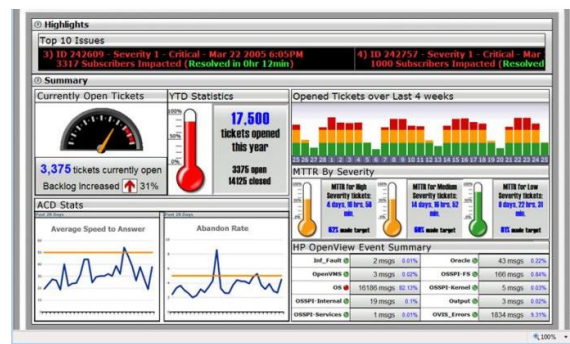
Drilling: Conceito de Agrupamento

	Qtde	Preco	PrecoCMS
Total de NmLinha	8.168	R\$ 7.187.324	R\$ 8.866.237
ELETRODOMESTICOS	2.329	R\$ 2.311.118	R\$ 2.819.590
ELETRODOMESTICOS, LINHA BRANCA	509	R\$ 367.588	R\$ 456.072
ELETRODOMESTICOS, LINHA BRANCA, FOGAO	220	R\$ 157.458	R\$ 196.835
ELETRODOMESTICOS, LINHA BRANCA, GELADEIRA	251	R\$ 181.139	R\$ 222.800
ELETRODOMESTICOS, LINHA BRANCA, LAVADORA	38	R\$ 28.990	R\$ 36.437
ELETRODOMESTICOS, ELETRODOMESTICOS DE COZINHA	1.076	R\$ 1.353.346	R\$ 1.639.292
ELETRODOMESTICOS, APARELHOS DE AR	744	R\$ 590.184	R\$ 724.227
TI	3.166	R\$ 2.605.301	R\$ 3.246.926
TI, MICRO-COMPUTADOR	451	R\$ 358.106	R\$ 447.132
TI, NETBOOK	1.347	R\$ 1.046.627	R\$ 1.305.692
TI, NETBOOK, XING-LING	338	R\$ 285.773	R\$ 354.906
TI, NETBOOK, BARATO	829	R\$ 600.584	R\$ 753.825
TI, NETBOOK, MARCA	180	R\$ 160.269	R\$ 196.961
TI, IPAD	913	R\$ 768.697	R\$ 955.943
TI, NOTEBOOK	455	R\$ 431.872	R\$ 538.159
SOM & IMAGEM	1.378	R\$ 1.145.022	R\$ 1.416.896
TELEFONIA	1.295	R\$ 1.125.882	R\$ 1.382.825

ARQUITETURA TÍPICA DE UM BI















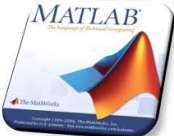







Administração do *Front End*

Apresentação correta para cada audiência



Treatments	β_a	β_{a1}	β_{a2}	β_{a3}	β_b	β_{b1}	β_{b2}	β_{b3}	R^2_t	F	MSD_{a1}	MSD_{a2}	MSD_{a3}	MSD_{b1}
6 x 6	8156.64	533.73	533.73	-7.26	1.2263	0.0724	0.0724	0.0816	91.21	1.19	0.09	0.09	1.70	-0.70
6 x 7	8332.58	523.73	868.60	33.80	1.3765**	0.0724	0.1373	0.1669	94.00	0.99	0.09	0.19	0.78	-0.07
6 x 11	7636.35	533.73	179.41	-173.24	1.0694	0.0724	0.0154	-0.0184	83.77	1.55	0.09	0.04	1.63	-0.21
6 x 21	7943.27	533.73	345.02	-31.94	0.9083	0.0724	0.0807	-0.2448*	77.29	1.99*	0.09	0.29	1.03	0.57
6 x 25	7661.85	533.73	632.96	664.63	0.7972	0.0724	-0.0972	-0.1780	53.13	4.59**	0.09	0.59	3.83	0.08
6 x 26	8639.35	533.73	330.40	678.77	1.2496*	0.0724	0.0273	0.1499	91.45	1.20	0.09	0.10	0.98	0.02
7 x 7	7784.99	868.60	868.60	-1048.66	0.8909*	0.1373	0.1373	-0.3837**	780.03	1.83	0.19	0.19	1.26	0.10
7 x 11	8315.15	868.60	179.41	370.70	1.2998*	0.1373	0.0154	0.1471	96.71	0.47	0.19	0.04	0.32	-0.08
7 x 21	9034.10	868.60	345.02	724.03	1.2410*	0.1373	0.0807	0.0229	82.41	2.69**	0.19	0.29	2.79	-0.58
7 x 25	7046.74	868.60	632.96	-263.35	0.7570**	0.1373	-0.0972	-0.2831*	69.24	2.08*	0.19	0.59	2.10	-0.79
7 x 26	8248.59	868.60	330.40	-46.86	1.1154	0.1373	0.0273	-0.0513	85.87	0.44	0.19	0.10	1.06	-0.91
11 x 11	6771.04	179.41	179.41	-684.23	1.0746	0.0154	0.0154	0.0438	74.23	3.28**	0.04	0.04	2.26	0.94
11 x 21	7764.14	179.41	345.02	140.26	0.8905	0.0154	0.0807	-0.2057	78.76	1.75	0.04	0.29	1.75	-0.33
11 x 25	6300.27	179.41	632.96	-134.63	1.0774	0.0154	-0.0972	0.1592	86.23	1.52	0.04	0.59	0.99	0.39
11 x 26	7919.63	179.41	330.40	313.36	0.9707	0.0154	0.0273	-0.0721	89.22	0.93	0.04	0.10	0.75	0.04
21 x 21	6670.08	345.02	345.02	-1116.42	0.9231	0.0807	0.0807	-0.2384*	67.88	3.38**	0.29	0.29	1.78	0.98
21 x 25	7303.41	345.02	632.96	664.59	1.1394	0.0807	-0.0972	0.1559	92.87	0.82	0.29	0.59	0.55	-0.61
21 x 26	8096.89	345.02	330.40	325.01	1.2756*	0.0807	0.0273	0.1676	89.62	1.54	0.29	0.10	1.50	-0.35
25 x 25	3865.33	632.96	632.96	1965.20	0.7678	-0.0972	-0.0972	-0.0378	66.76	2.40*	0.59	0.59	0.93	0.30
25 x 26	6727.95	632.96	330.40	-65.95	1.1015	-0.0972	0.0273	0.1714	91.80	0.89	0.59	0.10	0.69	-0.48
26 x 26	7613.47	330.40	330.40	-143.79	0.8996	0.0273	0.0273	-0.1551	78.92	1.77	0.10	0.10	1.31	0.27

BI: FERRAMENTAS DE MERCADO

ETL	DATA BASES	ADVANCED ANALYTICS	DATA DISCOVERY OLAP
     	    	   	      

QUADRANTE MÁGICO: BI E DW

Figure 1. Magic Quadrant for Business Intelligence and Analytics Platforms



Source: Gartner (February 2014)

Figure 1. Magic Quadrant for Data Warehouse Database Management Systems



<http://www.informationweek.com/big-data/big-data-analytics/gartner-bi-magic-quadrant-winners-and-losers/d/d-id/1114013>

<http://www.gartner.com/technology/reprints.do?id=1-1RP452A&ct=140310&st=sb>

Copyright © 2015 Prof. Jorge Surian

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).