

Especialização em Ciência de Dados com Big Data, BI e *Data Analytics*

Fundamentos de Business Intelligence (2° Encontro)

Prof. MSc. Fernando Siqueira



Sumário

- Visão Geral
- Data Warehouse
- Data Mart
- Business Intelligence
- Self-service Bl
- Ferramentas de Visualização de Dados
- Logical Data Warehouse



Representação dos Dados no Data Warehouse



1

Modelo de Dados de um DW

Modelo Dimensional

 Um modelo de dados baseado que suporta acesso de consulta de alto volume

Star schema

- Tipo de modelo dimensional mais comumente usado e o mais simples de modelagem.
- Contém uma tabela de fato conectada a várias tabelas de dimensão

Snowflakes schema

Uma extensão do esquema em estrela (star schema) onde o diagrama se assemelha a um floco de neve.



Modelo Entidade Relacionamento

Objetivo

- Eliminar, ao máximo, a redundância de dados
 - Permite que uma transação que promova mudanças no banco de dados atue o mais pontualmente possível

Necessário

- Fragmentar os dados por diversas tabelas
 - Traz uma considerável complexidade à formulação de consultas por um usuário final

100

Modelo Entidade Relacionamento

- Problemas do modelo ER para processamento analítico
 - Excessiva complexidade de representatividade gráfica do modelo → reduz a visão global do negócio
 - Usuário tem dificuldades em entender o modelo
 - Não apresenta alto desempenho na recuperação de dados, principalmente em operações de junção
 - A cada variação na estrutura do modelo → reescrever e ajustar as implementações

М

Modelo Entidade Relacionamento

- Problemas do modelo ER para processamento analítico
 - Questões típicas de análise dos negócios de uma empresa geralmente requerem a visualização dos dados segundo diferentes perspectivas
 - Exemplo: agência de automóveis que esteja querendo melhorar o desempenho de seu negócio, necessita examinar os dados sobre as vendas disponíveis na empresa
 - Uma avaliação deste tipo requer uma visão histórica do volume de vendas sob múltiplas perspectivas



- Multidimensional
 - Utiliza dimensões para armazenar e/ou visualizar os dados armazenados em um BD.

- Modelagem Multidimensional
 - Gera um modelo de dados simples de ser utilizado, principalmente por usuários que não são profissionais de informática
 - Fundamental para poder se trabalhar com ferramentas OLAP

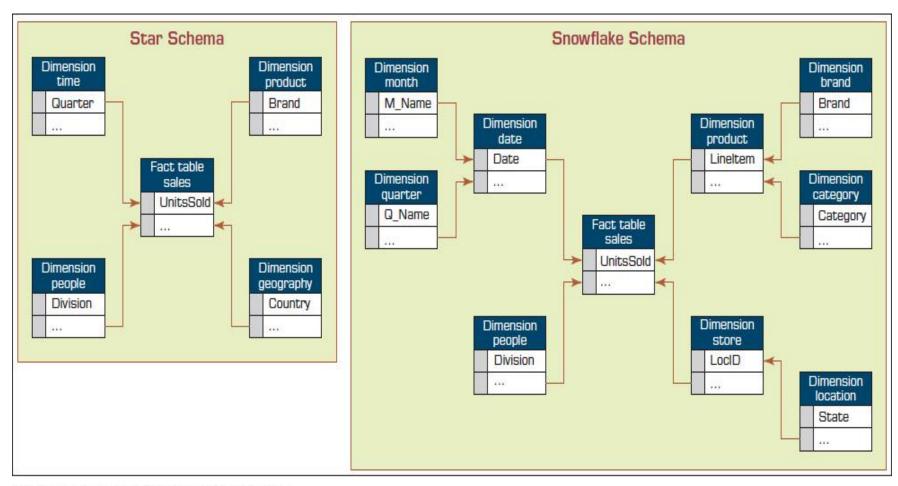
М

- Perspectivas
 - mês
 - modelo
 - loja
 - fabricante

- Visão histórica do volume de vendas sob múltiplas perspectivas:
 - volume de vendas por modelo
 - volume de vendas por loja
 - volume de vendas por período de tempo
 - volume de vendas por fabricante

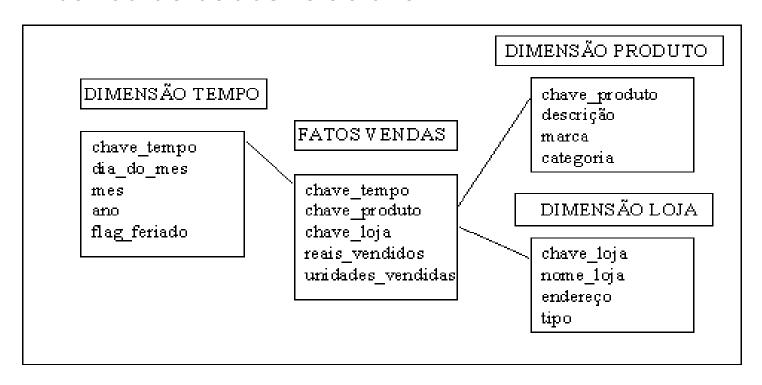


- Tabelas fato
 - Armazenam medições numéricas do negócio
 - Ex. valores, qtd. de ocorrências
- Tabelas dimensão
 - Armazenam as descrições textuais das dimensões do negócio
 - Ex. produto: identificador e nome





- Modelo Estrela Star Schema
 - Representação de um modelo dimensional em um banco de dados relacional





- Modelo Estrela Star Schema
 - A Modelagem Snowflake não é recomendada, pois dificulta o entendimento do modelo dimensional por parte do usuário e resulta em decréscimo de performance porque mais tabelas precisam ser unidas para satisfazer as consultas.



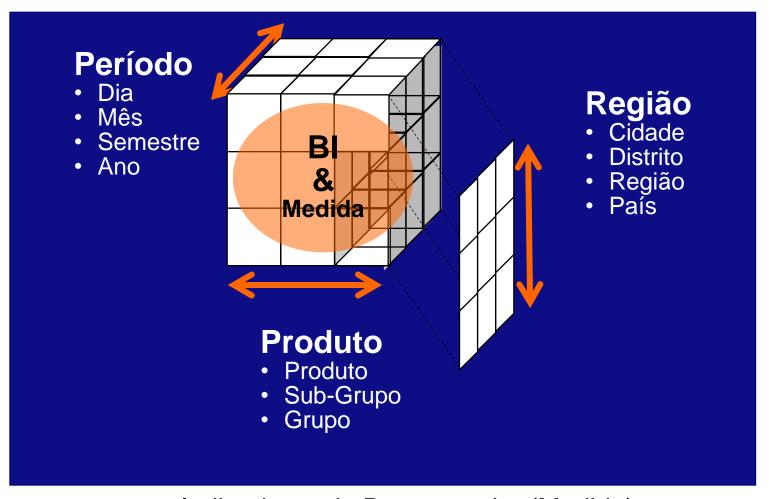
Modelo Estrela – Star Schema

 Oferece a capacidade de organizar, apresentar e analisar dados por várias dimensões, como vendas por região, por produto, por vendedor e por tempo

(quatro dimensões)

Dimens time					Dimension product
Quar	ter	1		ſ	Brand
		F	act tab sales UnitsS		
Dimens peopl	Contract Con	→	sales		Dimension peography

Análise Dimensional



Indicadores de Desempenho (Medida) Dimensões (Níveis de Análise)



Principais Características

- Um DW de pequena capacidade, projetado para atender uma área de negócio específica
- Dados mantidos no DW são separados por assunto, em subconjuntos, de acordo com:
 - Estrutura interna da empresa
 - O processo de tomada de decisão







Data Mart produção



Principais Características

- Pode desempenhar o papel de um DW departamental, regional, de acordo com:
 - Estrutura interna da empresa
 - O processo de tomada de decisão





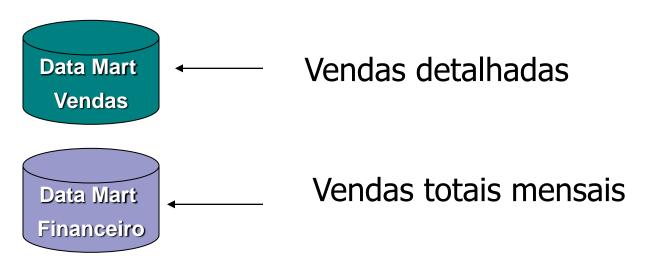
Data Mart Europa



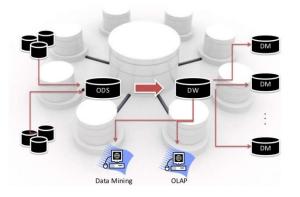


Principais Características

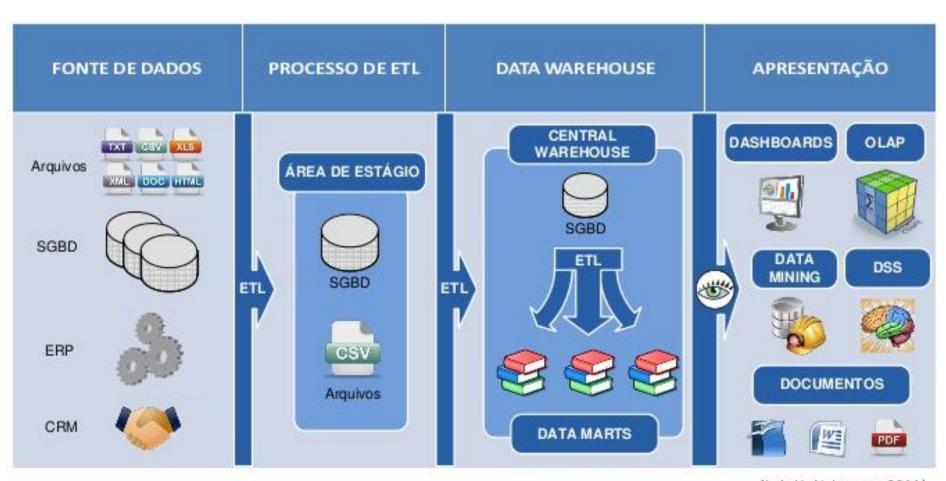
- Dados podem ser repetidos em dois ou mais Data Marts
- Os mesmos dados podem estar representados com granularidade diferente



Projeto de Data Warehouse



Arquitetura Genérica da Plataforma Bl



M

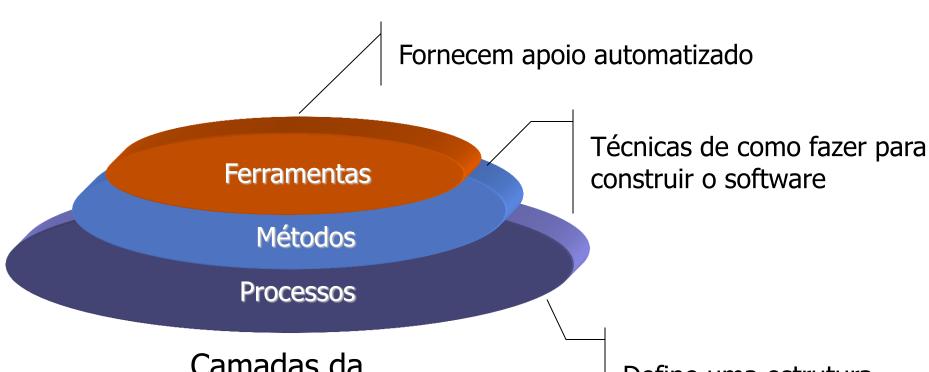
Processo de Desenvolvimento de Software

- O que é?
 - Uma série de passos, um roteiro, que se segue para criar, "a tempo" um resultado de alta qualidade
 - Resultado → Produto ou Sistema

- Por que é importante?
 - Fornece estabilidade, controle e organização



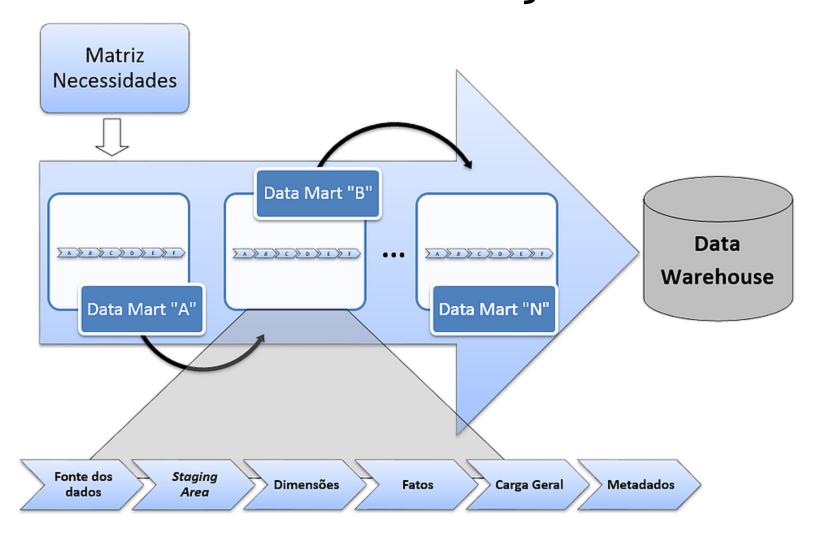
Processo de Desenvolvimento de Software



Camadas da Engenharia de Software

Define uma estrutura para efetiva utilização das técnicas e ferramentas

Processo de Construção do DW





Matriz de Necessidades

Identificar Métricas e Dimensões

 O objetivo desta atividade é gerar uma definição de alto nível dos indicadores que devem ser acompanhados pelo cliente, juntamente com as formas de visualização (dimensões) dos mesmos. Esse documento deve ser avaliado e aceito pelo solicitante do projeto

• Etapas:

- Analisar Relatórios e Gráficos
- Identificar Indicadores
- Analisar Demandas provenientes de outras áreas
- Definir Matriz de Indicadores e Dimensões

Matriz de Necessidades

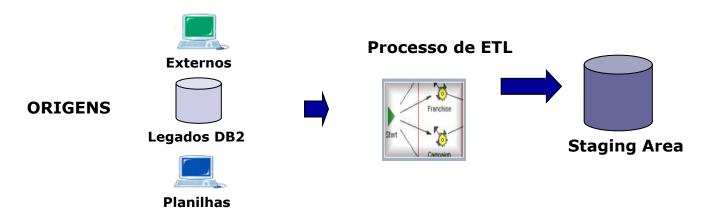
Dimensões Indicadores	Ano	Mes	Grupo Funcional	Diretoria	Unidade Solicitante	Legalidade	Natureza	Priorização	Sistema					
Demandas em Estoque	х	Х	X	Х	Х	X	X	X					\Box	\neg
Demandas Recebidas		х	х	х	х	х	х	х						\neg
Demandas Atendidas		Х	х	х	х	х	х	х						\neg
Demandas Atendidas no Prazo	Х	х	х	х	х	х	х	х						\neg
Percentual Atendimento no Prazo	Х	Х	Х	Х	Х	Х	Х	Х						\neg
Meta Atendimento no Prazo	Х	Х	Х	Х	Х	Х	Х	Х						
Demandas Reportando Defeitos		Х	Х	Х	Х	Х	Х	Х						
Percentual Reportando Defeitos		Х	Х	Х	Х	Х	Х	Х						
Meta Reportando Defeitos		Х	Х	Х	Х	Х	Х	Х						
Demandas Aguardando Aceite		Х	Х	Х	Х	Х	Х	Х						
Demandas Aceitas		Х	Х	Х	Х	Х	Х	Х						
Esforço Demandas Atendidas	Х	Х	Х	Х	Х	Х	Х	Х						

ĸ,

Construção SA, ODS, DM/DW

- Staging Area
- Operational Data Store
- Data Mart/Data Warehouse

SA - Staging Area



- Conhecida também com "camada integração e transformação" - Ilmon.
- Características:
 - Padronização
 - Tratamento dos dados
 - Transformações

ODS - Operational Data Store



- Visão integrada do mundo Operacional.
- Características:
 - Orientada a assunto
 - Coletivamente Integrada
 - Volátil
 - Valores correntes
 - Apenas dados comparativos detalhados

r

Construção SA, ODS, DM/DW

Projeto de Banco de Dados BI

- O objetivo desta atividade é projetar os modelos de dados do projeto.
- No contexto do BI são utilizados os seguintes tipos de Modelos de Dados:
 - Modelo de Dados do ODS (Operational Data Store) que contém as tabelas utilizadas no processo de recuperação de informações e na carga dos data marts.
 - Modelo de Dados Dimensional que contém as tabelas de dimensão e fato dos data marts.
 - Modelo de Dados da Stage Area que contém as tabelas da stage área usadas no processo de ETL.
 - Modelo de Dados do DW que contém as tabelas do Data Warehouse.

Etapas:

- Projeto de Banco de Dados
- Definir Volumetria

М

Construção SA, ODS, DM/DW

Etapas

- Projeto de Banco de Dados
 - A finalidade dos bancos de dados dimensionais é fornecer subsídio para realização de análises. Para tanto, sua arquitetura e até mesmo a terminologia empregada são distintas das utilizadas para bancos de dados transacionais.
 - Quando o modelo de dados começa a ser definido, elementos básicos de representação precisam ter sido estabelecidos, de modo a criar-se um padrão de modelagem. Em um modelo dimensional teremos dimensões e fatos representados em tabelas.

Definir Volumetria

- A análise de volumetria auxilia nas decisões de: design, arquitetura, modelagem,
 armazenamento, proteção de dados, planos de backup, latência, tráfego de rede, etc.
- A estimativa de volumetria deve ser aplicada aos modelos de Stage, Dimensional e ODS, deve contemplar o tamanho inicial da base de dados e a previsão de crescimento nos anos seguintes.

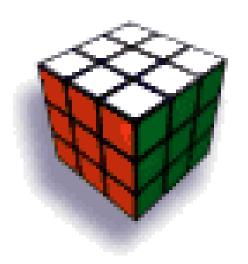


Modelagem Dimensional

- É uma técnica de projeto lógico que procura apresentar dados em uma forma comum que é intuitiva e permita acesso de alto desempenho
- Técnica antiga para criar BD simples e compreensíveis
- Modelo manipulado pelas ferramentas OLAP
- Modelo para a representação de assuntos passíveis de análise
- Contém as mesmas informações do MER (entidades, atributos e relacionamentos)



- Os dados são visualizados através de uma estrutura simples de "cubo de dados"
 - Visão mais simplificada do domínio
 - Facilmente extensível
 - Aumenta desempenho
 - Também chamado de "star schema"





Modelagem Dimensional

- Modelo manipulado pelas ferramentas OLAP
- Modelo para a representação de assuntos passíveis de análise
- Busca apresentar os dados em uma estrutura padronizada que é intuitiva e permite alto desempenho

Modelagem Dimensional



М

Modelagem Dimensional

Fatos

- São observações do negócio
- Um foco de interesse da empresa
 - Exemplos (domínio Loja):
 - Vendas, Promoções, Compras de fornecedores, ...
- Representação de um assunto
- Um assunto pode ser
 - Dado operacional, transação do negócio ou evento

M

Modelagem Dimensional

Dimensões

- Representação de contextos relevantes para a análise de um fato
- Granularidade adotada para representar fatos
- Dimensões exemplos para o Fato Vendas:
 - Clientes, Produtos, Tempo, Locais,
- Uma dimensão pode:
 - Conter membros
 - Ser organizada em hierarquias



Medidas

- Representação de atributos (variáveis) relevantes para a análise de um fato
- São normalmente valoradas e aditivas
- Indicadores de desempenho para análise
- Praticamente todas as consultas são construídas através de adições das medições



Medidas

São três tipos:

- Aditivas
 - Medidas que podem ser sumarizadas independente das dimensões utilizadas. Pode-se somar ao longo de qualquer dimensão
- Semi-aditivas
 - Medidas que podem ser sumarizadas em alguns casos. Isso porque a depender da situação empregada à métrica, pode-se perder sentido para a análise caso seja agregada. Isto é, permite ser somada ao longo de <u>algumas dimensões</u>.
- Não aditivas
 - Medidas que não podem ser sumarizadas ao longo das dimensões. Essas métricas não podem ter agregações, pois perdem a veracidade do valor.

M

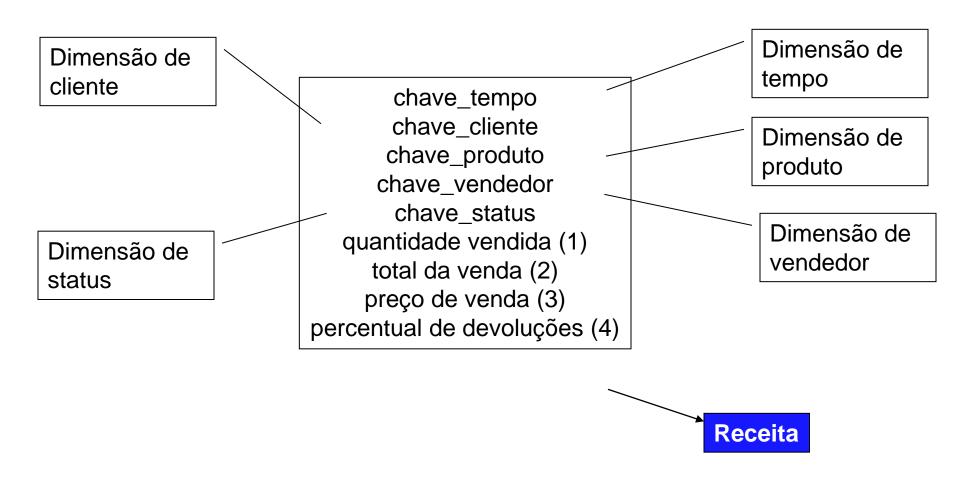
Modelagem Dimensional

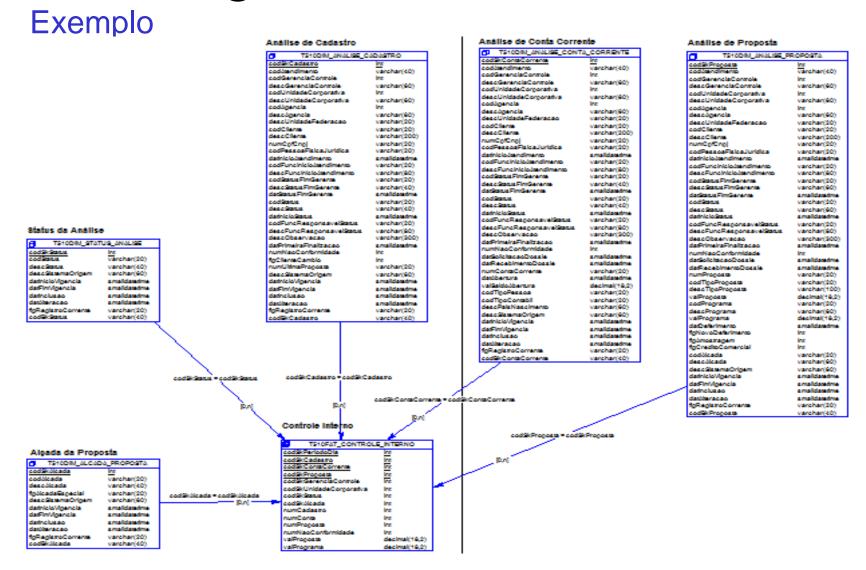
Medidas

- Uma medida é determinada pela associação de dimensões
 - Produto X Tempo.ano
 - Local.estado X Cliente.classe X Tempo.mês
- Exemplo para o Fato Vendas
 - Quantidade vendida por ano e produto;
 - Valor venda por estado e classe de cliente;
 - Percentual de devoluções por mês e vendedor.



Exemplo





Linhas: Colunas: ☐ Tipo de Pessoa ▼	Filtro de contexto: Ativa ▼		
Quantidade de Contas	2010/Jan	2010/Fev	2010
Conta Simplificada Pessoa Física	560.582	563.599	563.599
Conta Normal Crediamigo Pessoa Física	494.416	506.928	506.928
Demais Casos de Pessoas Físicas	128,244	129.464	129.464
Pessoa Jurídica - Cheque Empresa - Comércio	82.068	82.804	82.804
Beneficiário do INSS	80.309	80.629	80.629
Pessoa Jurídica para os Demais Tipos Contábeis Maiores que 04	28,200	28.665	28.665
Pessoa Física para os Tipos Contábeis 11, 20, 21, 22, 24, 29, 31, 32,	21.279	21.351	21.351
Funcionário do Banco do Nordeste	11.239	11.090	11.090
Conta Padrão Pessoa Física - Tomadores de Crédito	10.811	11.065	11.065
Funcionário Aposentado do Banco do Nordeste	6.198	6.112	6.112
Cliente que Recebe seus Vencimentos Através de Crédito em Conta	4.793	4.729	4.729
Pessoa Jurídica Optante do Simples - Cheque Empresa - Comércio	2.942	2.966	2.966
Mais			
Tipo de Pessoa	1.440.834	1.459.207	1.459.207

Medida, Indicador e Indicador Chave de Desempenho

Qual a diferença entre essas três palavras?

Medida: Dado que isolado não tem significado relevante, qualquer formalização de eventos observados.

Ex.: 50 litros de combustível

Indicador: Medida aplicada dentro de um determinado contexto de negócio. Ajuda a entender o que está acontecendo. Responde a uma pergunta específica. A sua forma de visualização tem um significado próprio.

Ex.: Medidor de combustível de um automóvel. Responde a pergunta: Quanto de combustível ainda posso utilizar? ou, Quanto de combustível consumi até agora?

Indicador Chave de Desempenho (KPI): É um indicador que nos ajuda a entender como estamos indo em relação a uma determinada meta. Ele nos leva à uma ação. Assim como o indicador, a sua forma de apresentação é parte crítica da sua definição.

Ex.: Sinalizador de reserva de combustível de um carro.





Melhores Práticas

- Todas as chaves do modelo devem ser chaves sem significado (Surrogate keys):
- Não se deve usar as chaves originais de produção (estas devem aparecer com atributos de dimensões).
- Um campo inteiro de quatro bytes pode conter mais de dois bilhões de valores, o suficiente para qualquer dimensão.
- Tabelas fatos não devem conter atributos. Apenas Chaves e Métricas



Melhores Práticas

- Não misture assuntos diferentes em uma mesma tabela fato
- Não misture granularidades diferentes em uma mesma tabela fato
- Na dimensão deve existir um único atributo que que se relaciona com uma tabela Fato. O relacionamento entre esse atributo e a fato é sempre um para n

Processo de Desenvolvimento do BI

subprocessos

Modelagem Dimensional

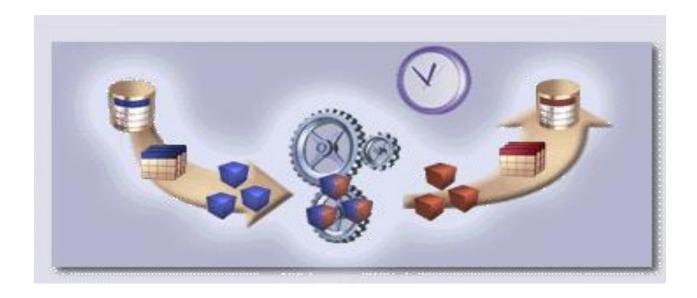
Extração e Transformação (fontes internas e externas

Construção dos DM's / DW

Desenvolvimento das Aplicações

ETL - Extract, Transform and Load

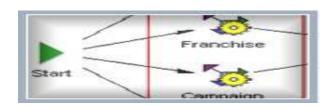
 Conjunto de técnicas para mover os dados de diversas fontes, de forma a promover a sua consolidação nas aplicações a que se destinam



м

ETL - Extract, Transform and Load

Processo de ETL



- É o processo de captura das fontes de dados a serem utilizadas em um ambiente BI, sua transformação, padronização e posterior carga no DW (ou DM ou ODS).
- Pode ser implementado via ferramenta específica ou através de linguagem de programação de banco de dados.
- Através de ferramentas são muitas as suas vantagens: metadados, facilidade de manutenção, gráfica, processo em uma só ferramenta, ...

ETL - Extract, Transform and Load

- Transformação
 - Após os dados serem extraídos, esses podem sofrer algumas transformações como:
 - □ Padronização das unidades, domínios, tipos de dados
 - □ "limpeza" dos dados
 - Combinação das diversas fontes de dados
 - Criação de surrogate keys para cada registro dimensional
 - Onde devem ser executadas as transformações:
 - Onde existe capacidade de processamento disponível
 - Nas plataformas que a organização está habituada a operar

ETL - Extract, Transform and Load

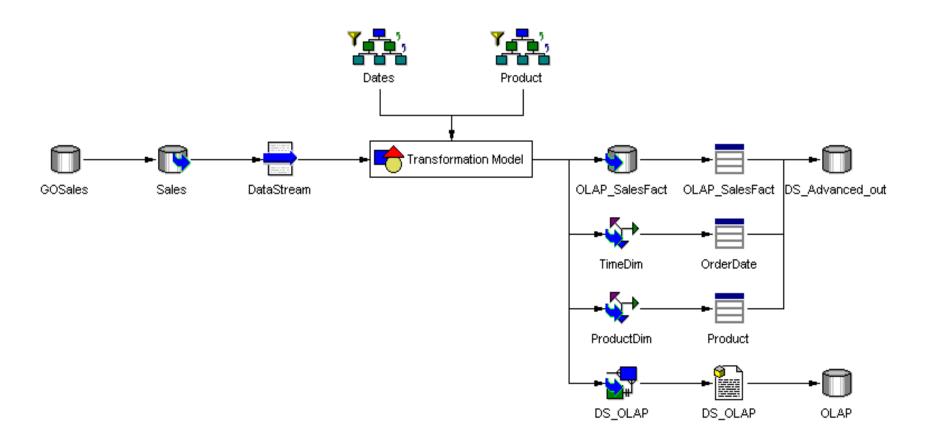
- Carga/População
 - Após o processo de transformação, os dados já estão aptos a serem carregados no DW
 - A carga consiste em povoar as tabelas do modelo dimensional (tabelas dimensões e fatos) criadas no repositório DW

ETL – Extract, Transform and Load

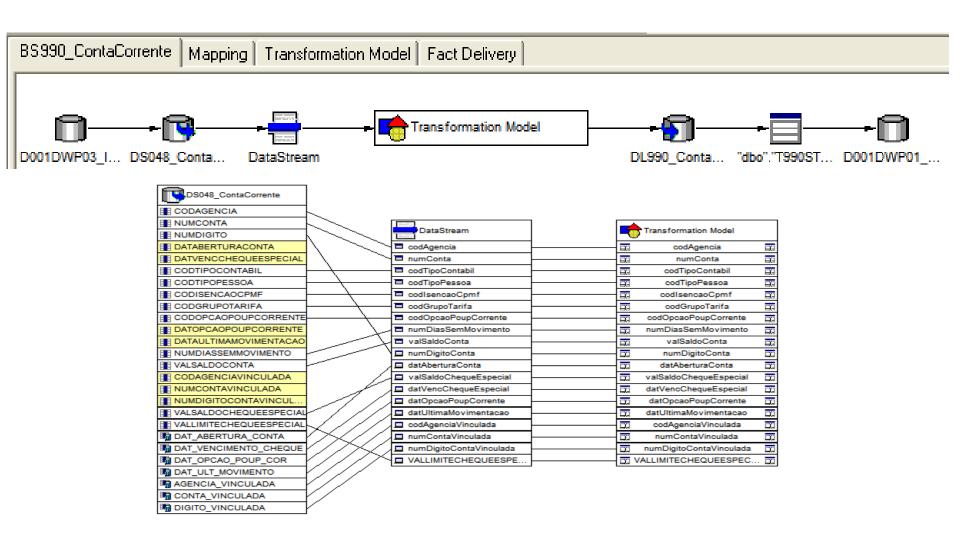
	Mapeamento					
Versão						
Nome do Job						
Nome da Transformação						
Origem da Informação						
Destino da Informação						
Observação						
Regra do Join						
Regra da Seleção						
ORIG			DESTINO			
Tabela	Atributos	Tabela	Atributos	Transformação / Valor		
		C				
		Consulta SQI	L			

×

ETL – Extract, Transform and Load

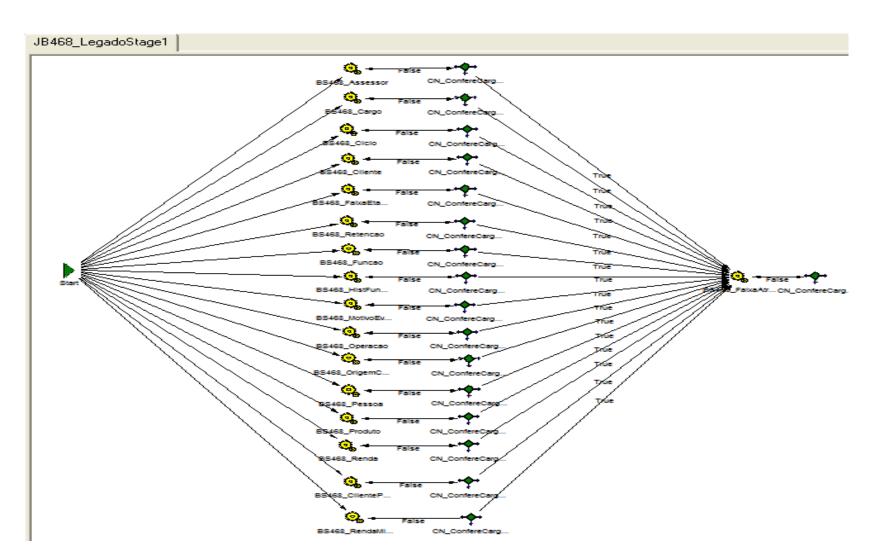


ETL - Extract, Transform and Load



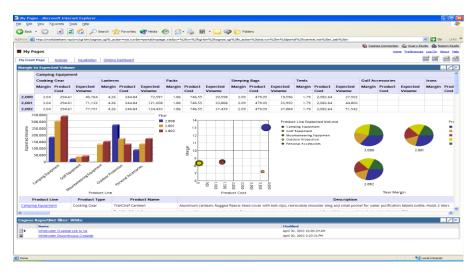
М

ETL – Extract, Transform and Load



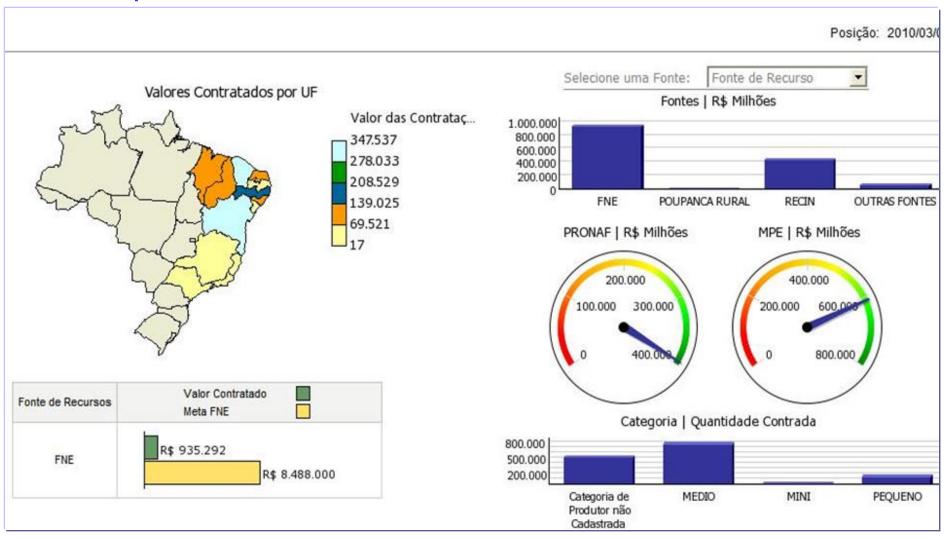


Conceito



- DASHBOARD Apresenta complexas informações de negócios de forma rápida e intuitiva. Com a visualização simultânea de diversos quadros possibilitanto analisar as informações que se correlacionam e, assim, tomar decisões melhores.
- Impacto visual instantâneo.
- Exploração Interativa. Analise com mais detalhes (Zoom) uma área de interesse, podendo visualizar diferentes novas perspectivas.

Exemplo





- Desing de Dashboard
 - O desafio fundamental do design do dashboard é exibir todas as informações necessárias em uma única tela, de forma clara e sem distrações, de forma que possa ser rapidamente assimilada
- As Camadas de Apresentação
 - Monitoramento
 - Análise
 - Gerenciamento



- O que procurar em um dashboard
 - Uso de componentes visuais para destacar dados e exceções que exigem ação
 - Transparente para o usuário, o que significa que requer treinamento mínimo e são extremamente fáceis de usar
 - Combinar dados de uma variedade de sistemas em uma visão única, resumida e unificada do negócio
 - Habilitar drill-down ou drill-through para fontes de dados ou relatórios subjacentes
 - Apresentar uma visão dinâmica do mundo real com dados atuais
 - Exigir pouca codificação para implementar, implantar e manter

Melhores Práticas de Design para Das

Melhores Práticas de Design para Dashboard

- KPI de referência com padrões da indústria
- Envolver as métricas com metadados contextuais
- Validar o Design por um Especialista em Usabilidade
- Priorizar e classificar alertas e exceções
- Enriquecer o painel de controle com comentários de empresas e usuários
- Apresentar informações em três níveis diferentes
- Fornecer análise orientada



Especialização em Ciência de Dados com Big Data, BI e *Data Analytics*



Prof. Fernando Siqueira

fssiqueira@gmail.com