

A decorative graphic on the left side of the slide consisting of a grid of squares in various shades of blue and purple, some of which are solid and others are semi-transparent, creating a layered effect.

Fundamentos de *Business Intelligence* (3º e 4º Encontros)

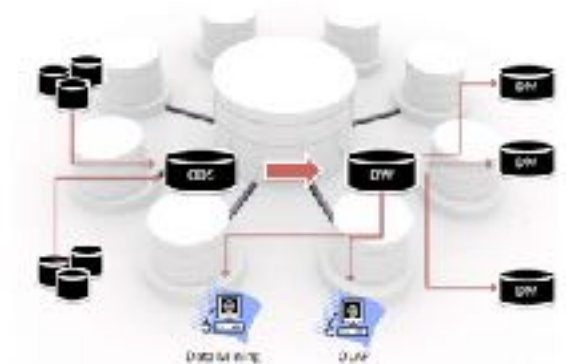
Prof. MSc. Fernando Siqueira

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms

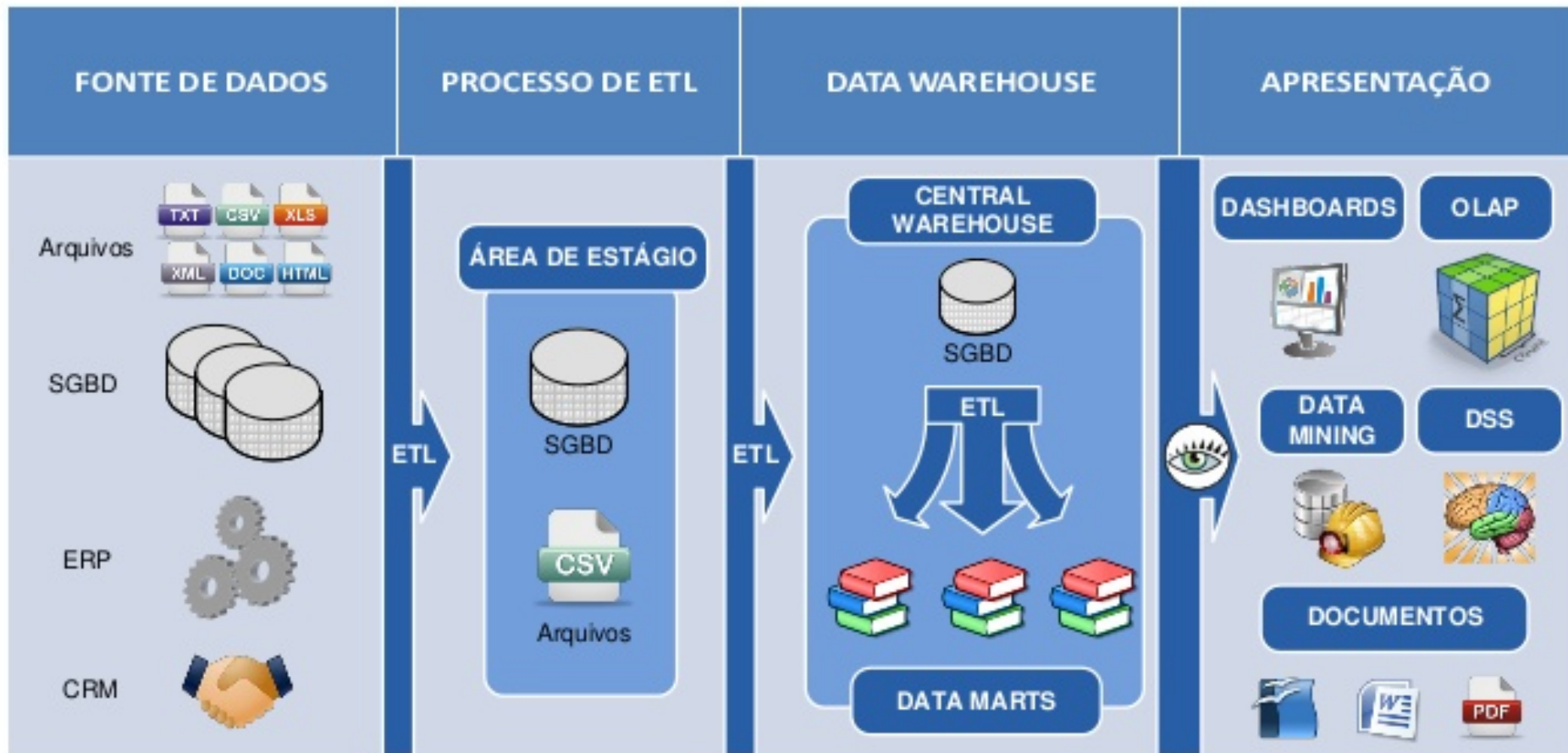


Source: Gartner (February 2010)

Projeto de *Data Warehouse*



Arquitetura Genérica da Plataforma BI

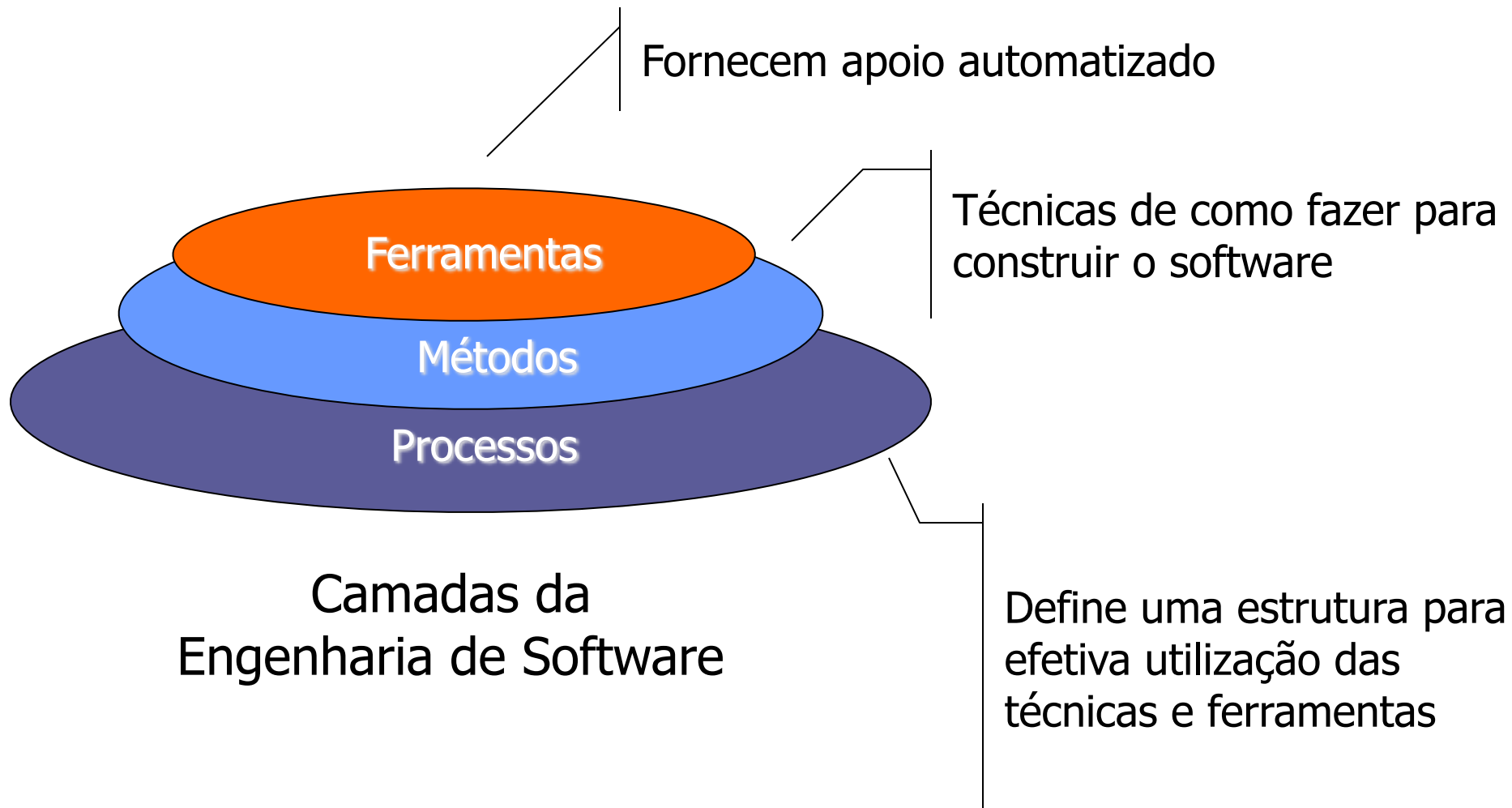




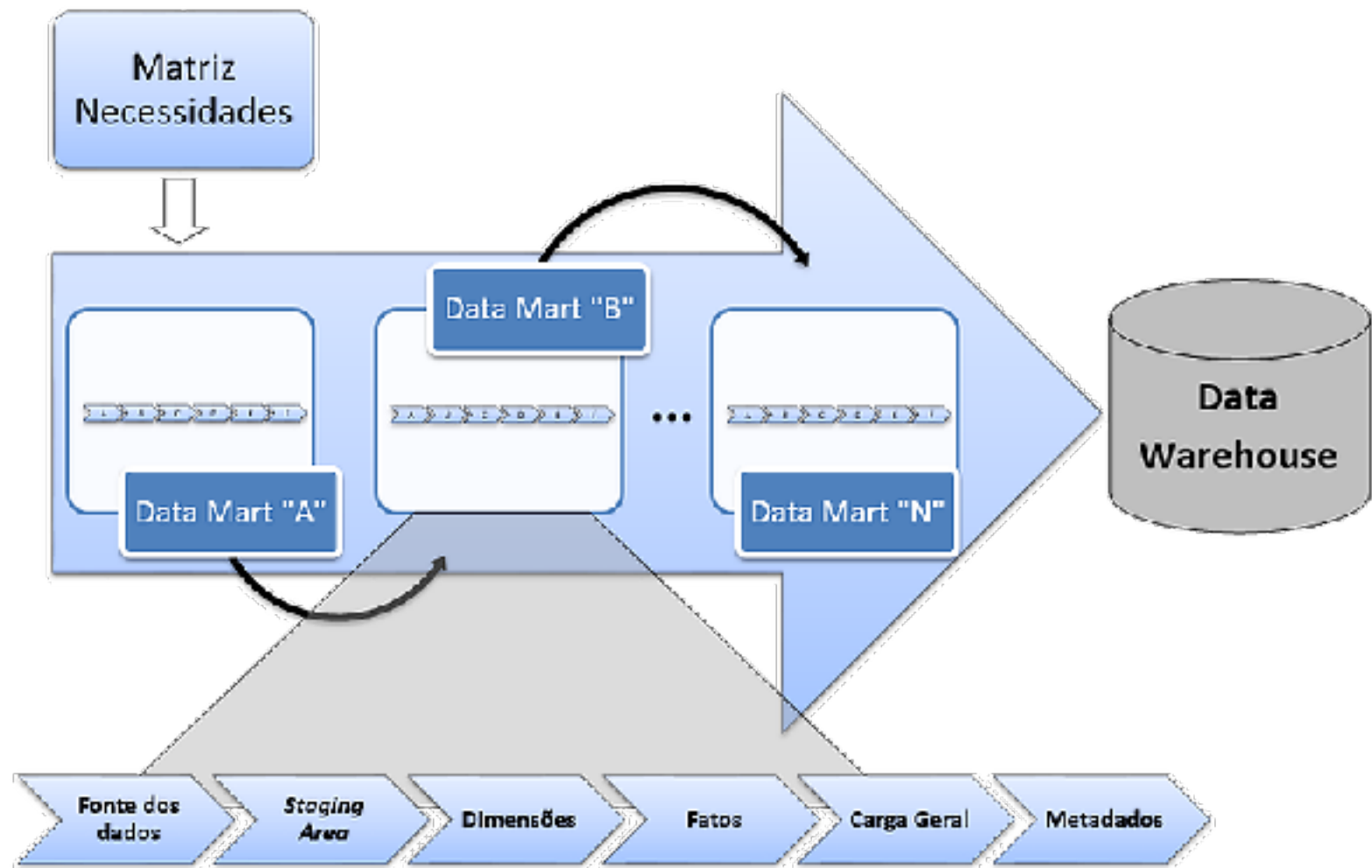
Processo de Desenvolvimento de Software

- 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



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

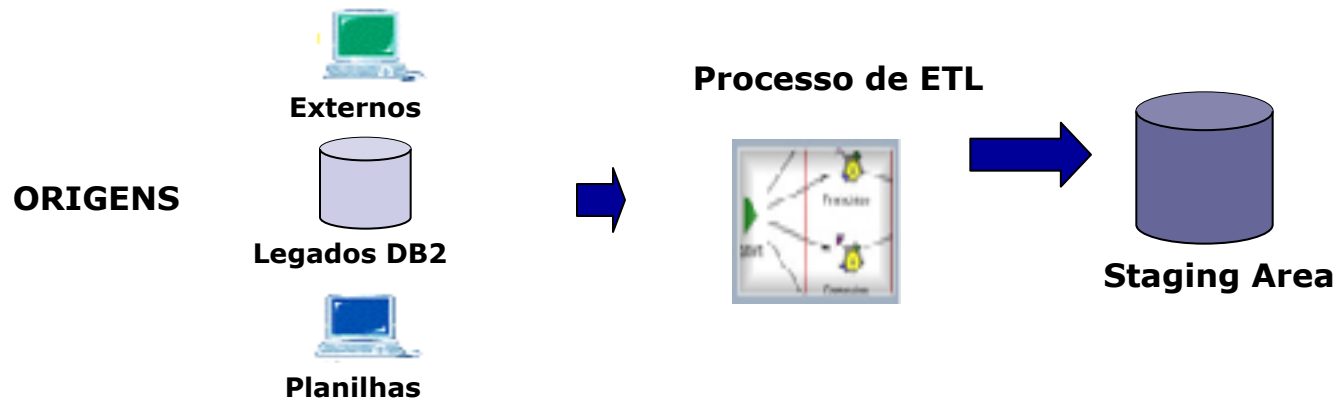
Matriz Necessidade	Dimensões	Sexo	Raça	Faixa Etária	Profissão	Estado	Cidade	Bairro	Tempo	
Indicadores										
Quantidade de Portadores de Deficiência		X	X	X	X	X	X	X	X	
Quantidade de Empregados		X	X	X	X	X	X	X	X	
Total da Renda de Empregados		X	X		X	X	X	X	X	
Total da Renda de Portadores de Deficiência		X	X		X	X	X	X	X	
% Desempregados		X	X	X		X	X		X	



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



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



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)

Modelagem Dimensional

- 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”

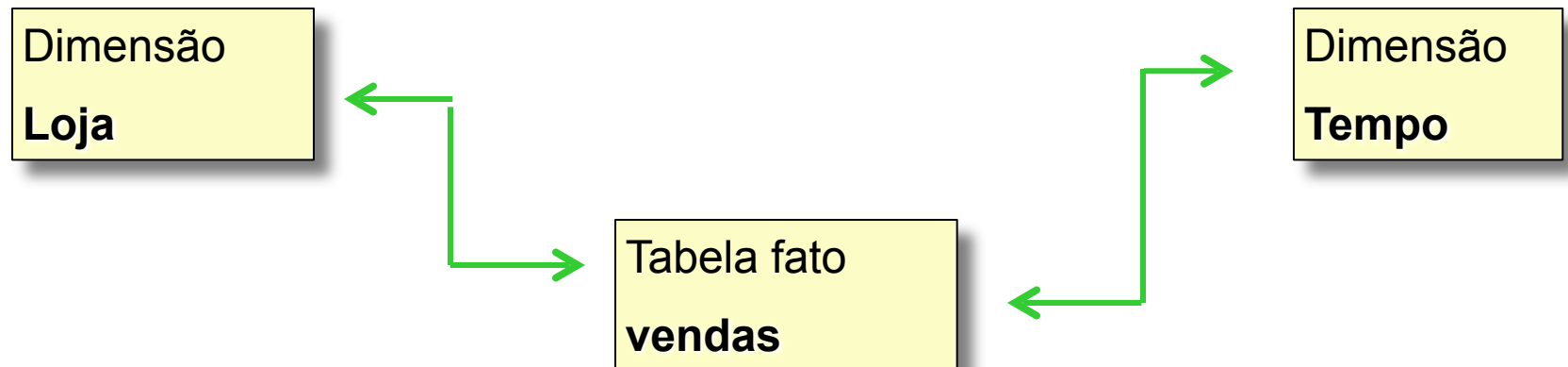




Modelagem Dimensional

- 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





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



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



Modelagem Dimensional

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

Modelagem Dimensional

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 algumas dimensões.
 - Não aditivas
 - Medidas que não podem ser sumarizadas ao longo das dimensões. Essas métricas não podem ter agregações, pois perdem a veracidade do valor.



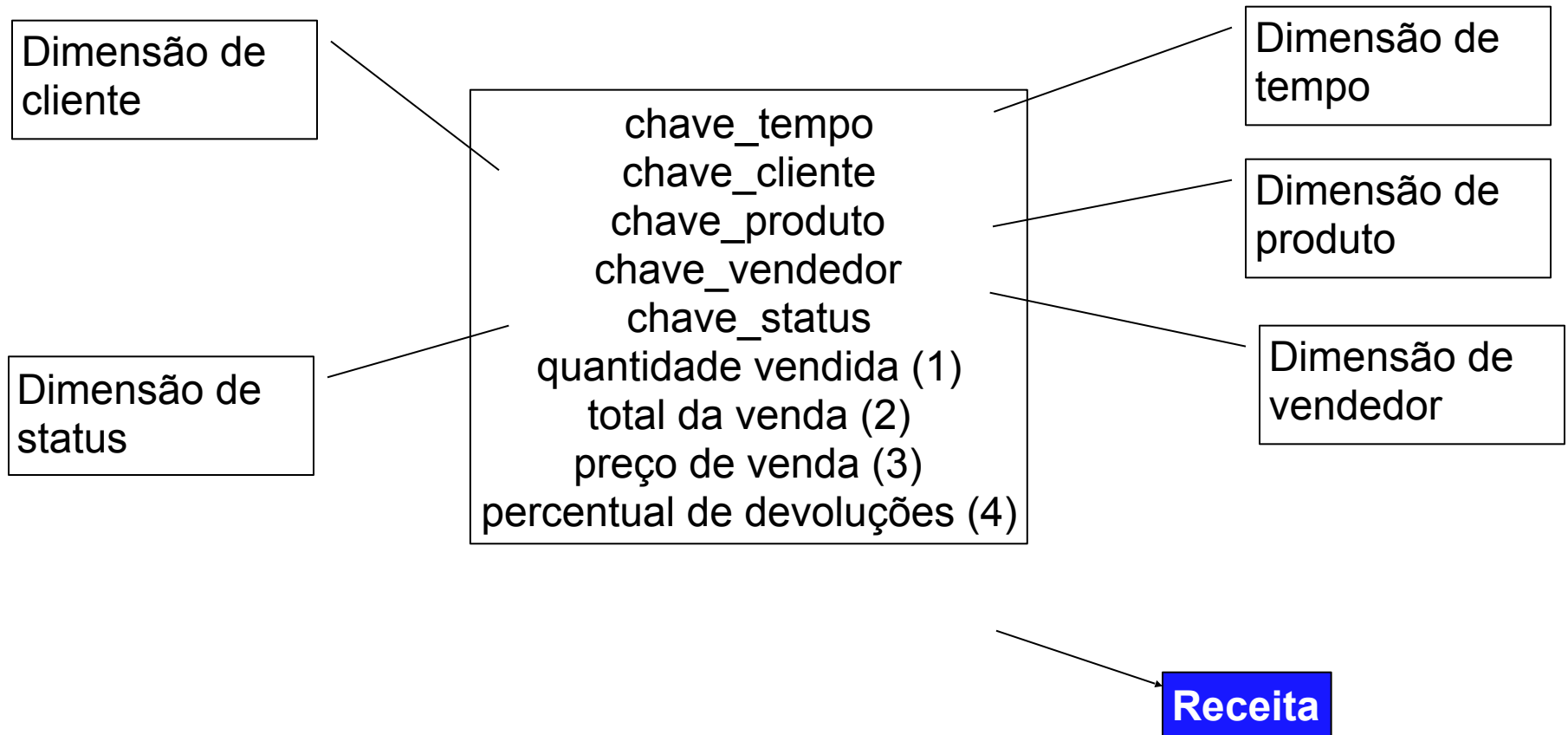
Modelagem Dimensional

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.

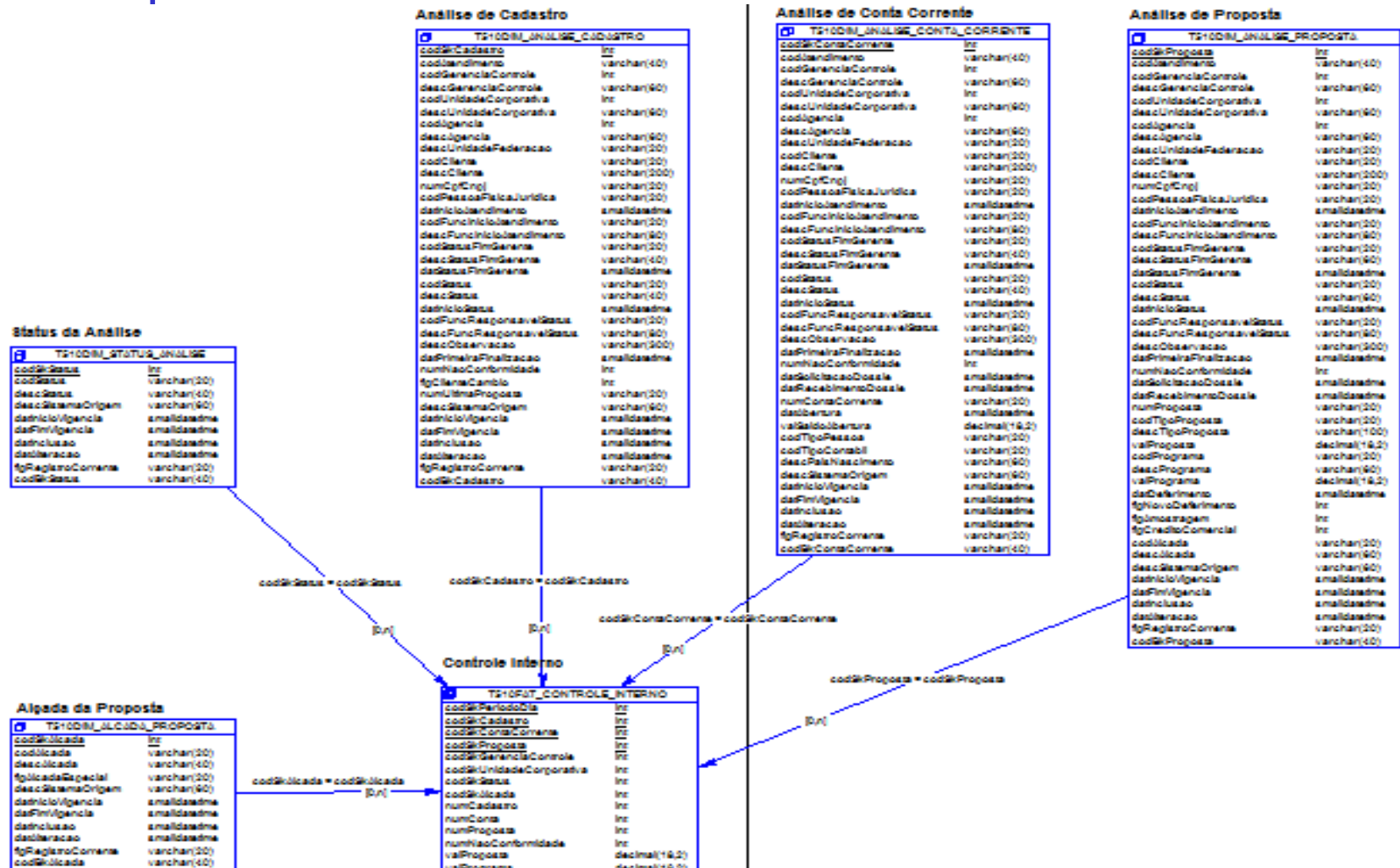
Modelagem Dimensional

Exemplo



Modelagem Dimensional

Exemplo



Modelagem Dimensional

<div> <div>Linhas:</div> <div> <div> <div> <div></div> <div>Tipo de Pessoa</div> <div></div> </div> </div> </div> <div>Colunas:</div> <div> <div> <div></div> <div>2010</div> <div></div> </div> </div> </div> <div>Filtro de contexto:</div> <div> <div> <div></div> <div>Ativa</div> <div></div> </div> </div>
--

Modelagem Dimensional

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.

Modelagem Dimensional





Modelagem Dimensional

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



Modelagem Dimensional

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 se relaciona com uma tabela Fato. O relacionamento entre esse atributo e a fato é sempre um para n



Modelagem Dimensional

As 10 Regras Essenciais para a Modelagem de Dados Dimensional by Ralph Kimball

1. Carregue dados detalhados para as estruturas dimensionais
2. Estructure os modelos dimensionais em torno dos processos de negócios
3. Tenha certeza de que cada tabela fato tenha uma dimensão de data associada.
4. Certifique-se que todos os fatos em uma única tabela fato estão na mesma granularidade ou nível de detalhe
5. Resolva relacionamentos muitos-para-muitos em tabelas fato



Modelagem Dimensional

As 10 Regras Essenciais para a Modelagem de Dados Dimensional by Ralph Kimball

6. Resolva os relacionamentos muitos-para-um nas tabelas de dimensões.
7. Gravar nomes de relatórios e valores de domínios de filtros em tabelas dimensão.
8. Tenha certeza de que as tabelas dimensão usam uma chave artificial.
9. Crie dimensões padronizadas para integrar os dados na empresa.
10. Avalie requisitos e realidade continuamente para desenvolver uma solução de DW/BI que seja aceita pelos usuários de negócios e suporte seu processo de tomada de decisões.

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



Prof. Fernando Siqueira
fernando.siqueira@uni7.edu.br