



Modelagem, Extração e Manipulação de Dados

BLOCO: B.I. E ANÁLISE DE DADOS

PROF. RODRIGO EIRAS, M.SC.

[ETAPA 1] AULAS 1 E 2 – MODELAGEM DIMENSIONAL



No último trimestre...

- Fundamentos de B.I.
- Modelagem Relacional
- Fundamentos de Big Data
- Análise de Dados em Big Data
- Power BI



Agenda

- Introdução à disciplina de Modelagem, Extração e Manipulação de Dados
- Modelagem de dados
- Modelagem Relacional x Dimensional
- Modelagem Dimensional
- Quiz

Introdução

- Nesta disciplina, você irá estudar modelagem de dados para Business Intelligence (BI) - modelagem dimensional - e compreender como obter e transformar dados em informação para tomada de decisão.
- Vai conhecer em mais detalhe a ferramenta Power BI da Microsoft enquanto entra em contato com as melhores práticas de apresentar a informação para tomada de decisão.

Introdução ao curso

Etapa 1: Modelagem
de Dados e
Modelagem
Dimensional

Etapa 2: Tipos de
Dados e Modelagem
Dimensional

Etapa 3: Conectar
Dados em Banco de
Dados Utilizando
MPBI

Etapa 4: Visualizar
Dados no Power BI

Introdução ao curso

Etapa 5: Analisar
Dados no Power BI -
parte 1

Etapa 6: Analisar
Dados no Power BI -
parte 2

Etapa 7: Extrair dados
de sistemas de
informação utilizando
SQL - parte 1

Etapa 8: Extrair dados
de sistemas de
informação utilizando
SQL - parte 2

Etapa 9: Elaborar
visualizações
analíticas: relatórios e
dashboard

Ao final desta disciplina, você será capaz de:

- Utilizar modelagem de dados;
- Utilizar o Microsoft Power BI (MPBI);
- Extrair dados de sistemas de informação utilizando SQL;
- Criar visualizações e cruzamentos de dados de sistemas de informação em uma plataforma de BI.

Introdução ao curso

A modelagem de dados possibilita o alcance de quatro objetivos:

- ajuda a visualizar o sistema como ele é ou como deve ser;
- permite especificar a estrutura ou o comportamento de um sistema;
- proporciona um guia para a construção do sistema e;
- documenta as decisões tomadas.

Fonte: (BOOCH; RUMBAUGH; JACOBSON, 2000).

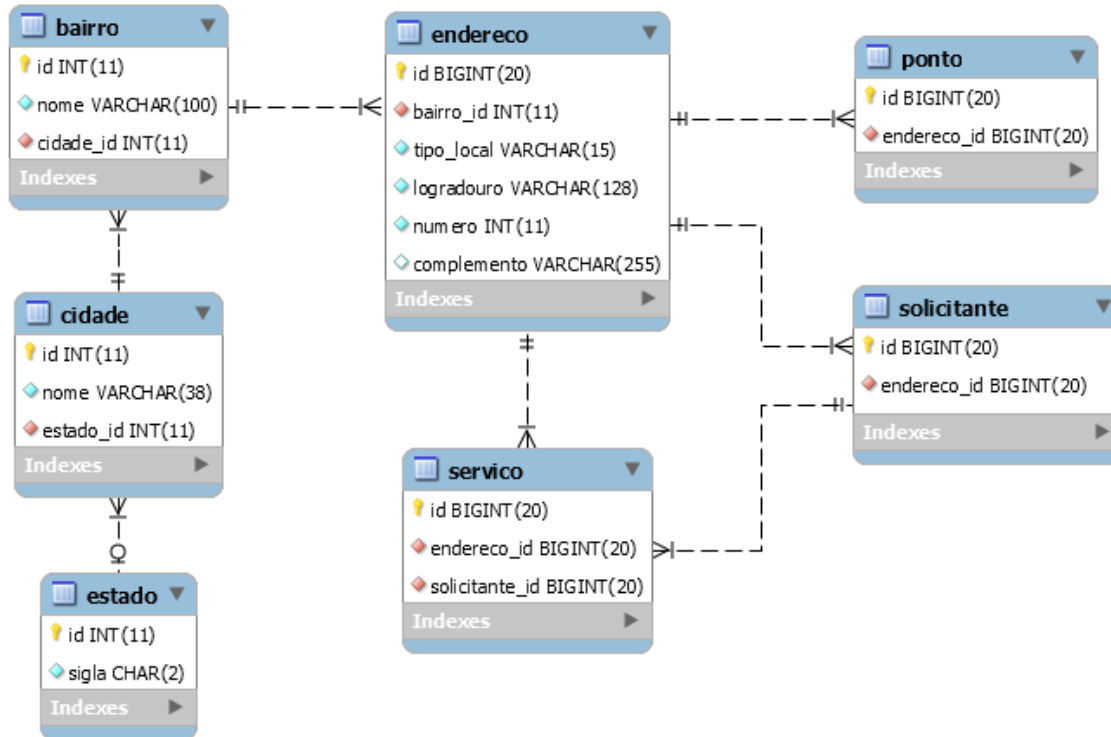
Modelagem de dados

Modelagem de dados

- A modelagem de dados é uma etapa estruturante da arquitetura de soluções de sistemas de informações.
- Em soluções de Business Intelligence (BI), a etapa de modelagem de dados é a mais importante, pois a principal entrega de uma solução de BI são dados e informações para a tomada de decisão.
- Uma modelagem de dados com alta qualidade em soluções de BI pode agregar valor ao negócio que extrapolam o objetivo da solução, promovendo insights para a transformação do negócio.
- Vamos fazer uma revisão na sequência

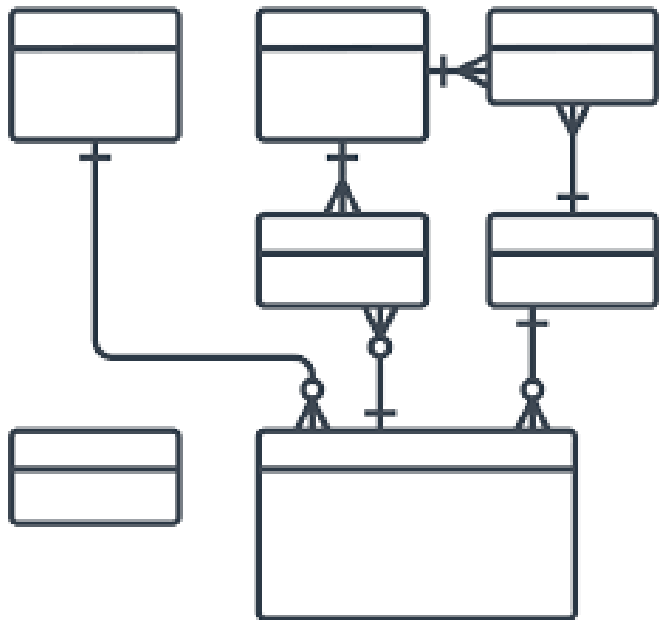


Modelagem Relacional vs Dimensional

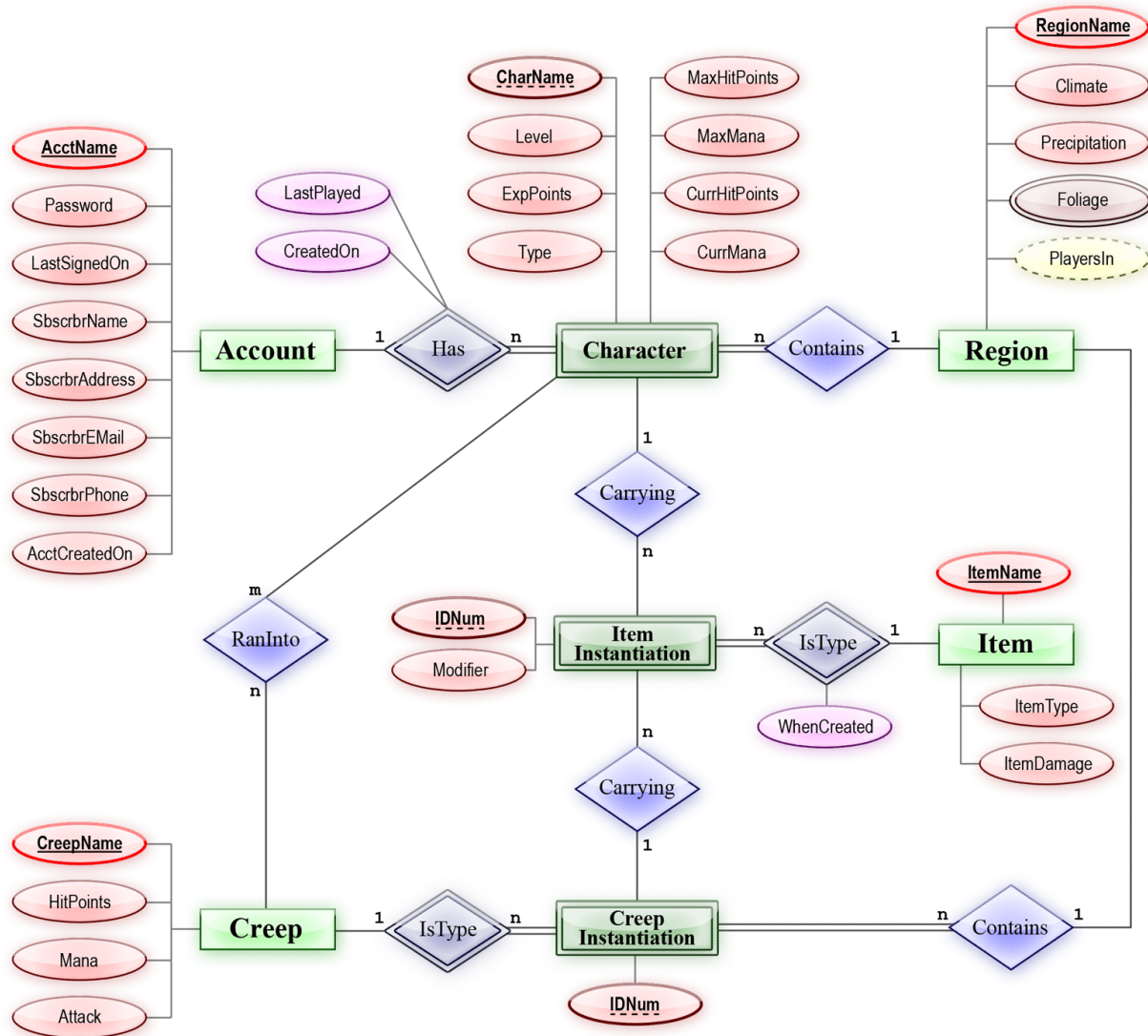


- Um dos modelos para modelagem de dados é o Modelo Relacional.
- É utilizado na modelagem de dados OLTP (*OnLine Transaction Processing*) e tem como objetivo um armazenamento consistente e eficiente ao dado.
- Essa forma de modelagem diminui a redundância de dados e as inconsistências, os dados estão normalizados na terceira forma normal.
- O foco é inserir, alterar e deletar os dados.

Modelo relacional!



- Fácil de usar
- Rigor matemático na representação de dados
- Estrutura de dados simplificada
- Ausência de detalhe de desempenho e implementações
- Informações são armazenadas em relações normalizadas

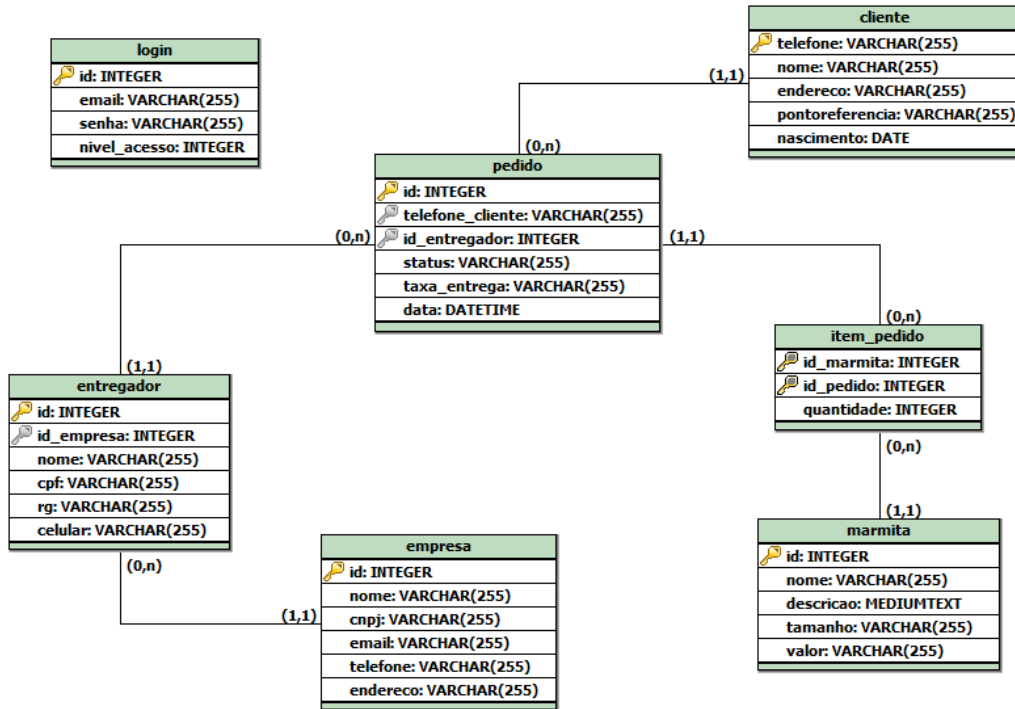


MER – Modelo Entidade Relacionamento

- É um modelo abstrato cuja finalidade é descrever, de maneira conceitual, os dados a serem utilizados em Sistema de Informação ou pertencente a um determinado contexto.

MER – Modelo Entidade Relacionamento

- A principal ferramenta do modelo é a sua representação gráfica, o DER.
- Objetivo:
 - Facilitar o projeto de banco de dados, possibilitando especificar a estrutura lógica geral do banco de dados.



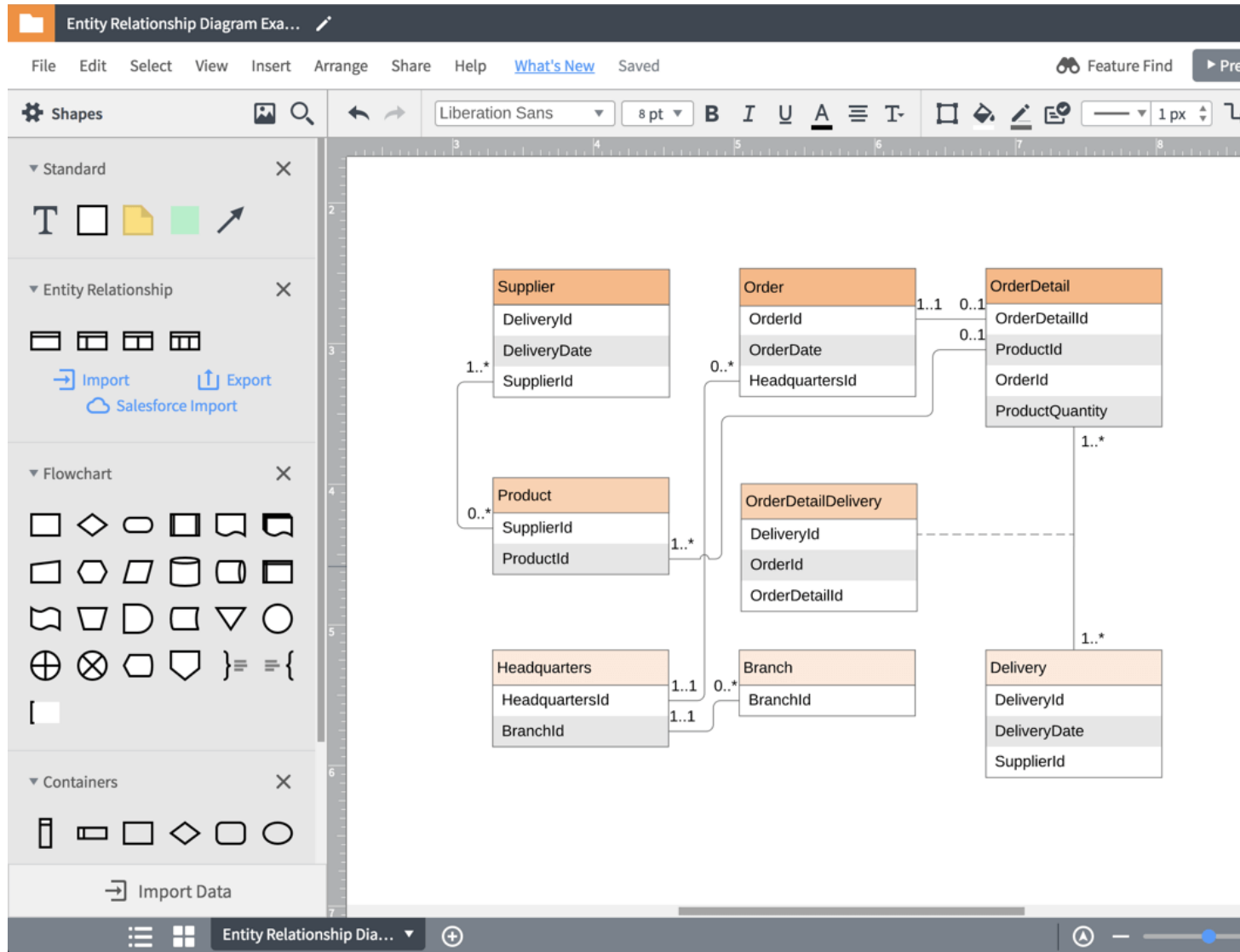


Diagrama Entidade Relacionamento

- Componentes
 - Entidade
 - Atributos
 - Relacionamentos

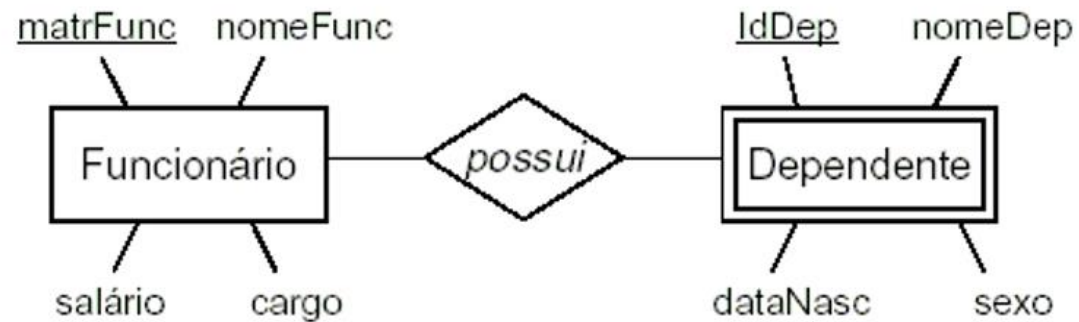
DER - Entidade

- É uma representação concreta ou abstrata de um objeto, com características semelhantes, do mundo real.
- Ex.: Fornecedor, Pessoa, Imóvel, Curso.

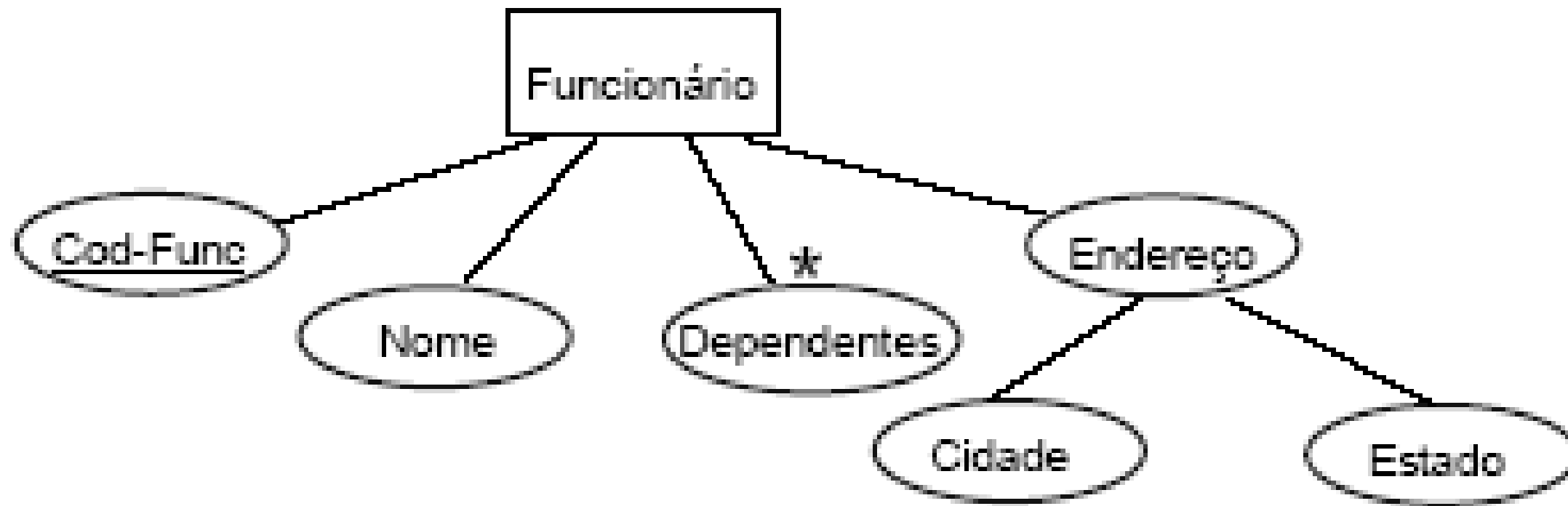


EMPREGADO

DER – Entidade Fraca



- Não existe se não estiver relacionada a outra, isto é, ela é logicamente dependente da outra.
- Alguns conjuntos entidade não possuem um conjunto de atributos capaz de identificar univocamente uma determinada entidade. Neste caso, sua existência depende da existência de outra entidade.



DER - Atributo

ELEMENTO DE DADO QUE CONTÉM O VALOR DE UMA PROPRIEDADE DE UMA ENTIDADE.

Atributo Simples

Não possui qualquer característica especial. A maioria dos atributos serão simples. Quando um atributo não é composto, recebe um valor único como nome, por exemplo e não é um atributo chave, então ele será atributo simples.

A maioria dos atributos são considerados simples.

Em uma entidade cliente, por exemplo, poderemos considerar como atributo simples: nome, sexo, data de nascimento, dentre outros.

DER – Atributo/Classificação

Atributo Composto

O seu conteúdo é formado por vários itens menores. Exemplo: Endereço. Seu conteúdo poderá ser dividido em vários outros atributos, como: Rua, Número, Complemento, Bairro, Cep e Cidade. Este tipo de atributo é — chamado de atributo composto. Veremos mais de sua aplicação no post sobre normalização de dados.

É importante considerar que na aplicação do banco de dados um atributo composto geralmente é desmembrado, ou seja, para o caso do endereço, podemos desmembrá-lo em vários atributos simples, como: Rua, número, complemento, bairro, cidade e cep. Conceitualmente é aceito o endereço como um único atributo, mas na prática geralmente é feito este desmembramento para permitir a organização dos dados inseridos e facilitar a busca e indexação dos mesmos.

DER – Atributo/Classificação

Atributo Multivalorado

O seu conteúdo é formado por mais de um valor.

Exemplo: Telefone. Uma pessoa poderá ter mais de um número de telefone. É indicado colocando-se um asterisco precedendo o nome do atributo. O atributo multivalorado serão tratados com mais detalhes na normalização de dados.

DER – Atributo/Classificação

Atributo Determinante

Identifica de forma única uma entidade, ou seja, não pode haver dados repetidos.

É indicado sublinhando-se o nome do atributo. Exemplo: CNPJ, CPF, Código do fornecedor, Número da matrícula, etc. Os atributos determinantes serão as chaves primárias no banco de dados e seu uso tem implicações na normalização de dados.

DER – Atributo/Classificação

DER – Considerações importantes relacionadas ao negócio

- A análise de atributos é parte importante da análise e modelagem de dados.
- A quantidade deles, tipo e outras informações a seu respeito geralmente permitirá a construção de um banco de dados com melhor performance.
- PERGUNTA IMPORTANTE:
 - Quem define os atributos, o cliente ou o analista de sistema?
 - A resposta é que deve ser um trabalho conjunto e levando em conta o bom senso. É preciso considerar que há muita subjetividade na análise de sistemas.
 - Normalmente o processo é conduzido por um analista, mas não cabe necessariamente a ele fazer todas as definições, uma vez que ele trabalha com dados fornecidos pelo cliente ou obtidos no ambiente do mesmo.

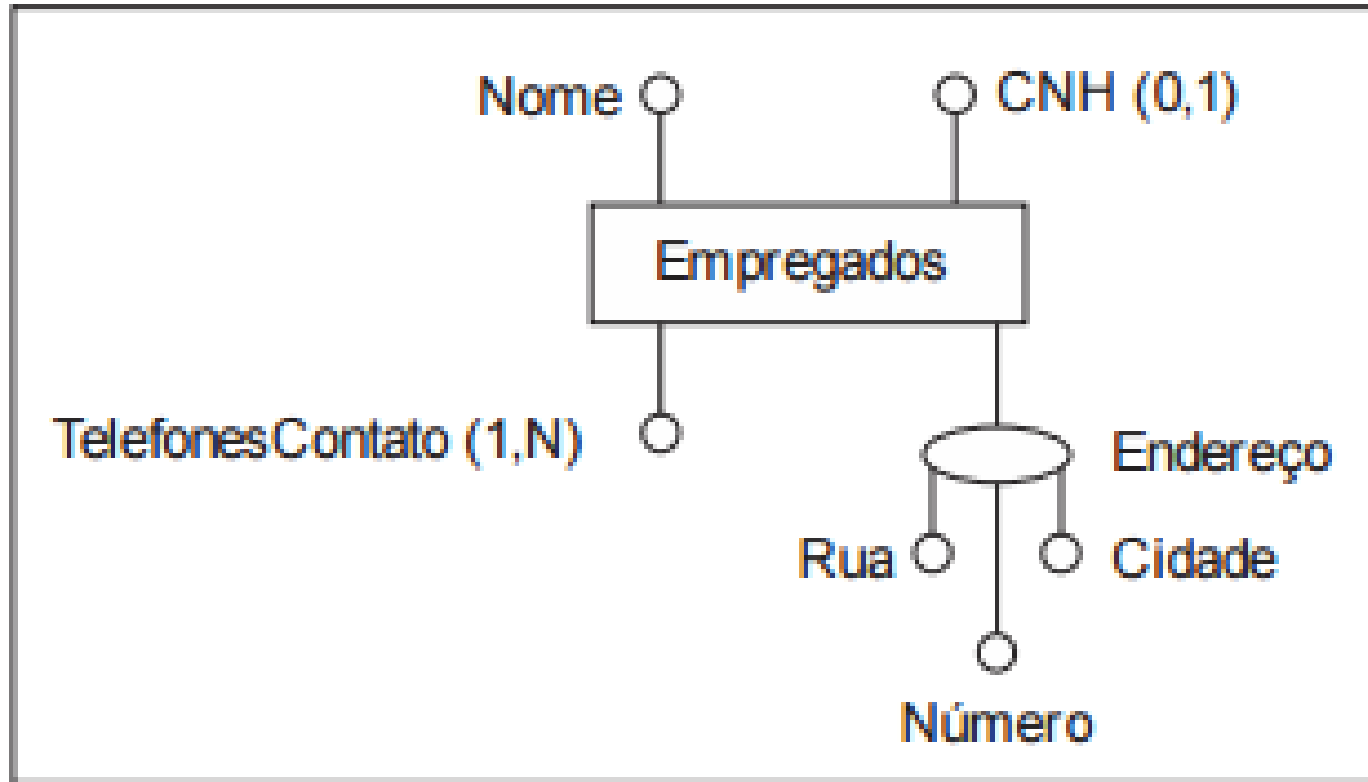




DER – Considerações importantes relacionadas ao negócio

- É preciso levar em conta que um sistema de informação não é um objeto estático, muito pelo contrário, ele é extremamente dinâmico e sofre modificações o tempo todo.
- Assim aquela ideia de o que você vai concluir um sistema um banco de dados, pode servir apenas para você definir um ponto onde quer chegar
- Mas quando chega esse ponto verá que outros pontos já estão sendo estabelecidos e portanto mudanças são inevitáveis.

Mais sobre atributos



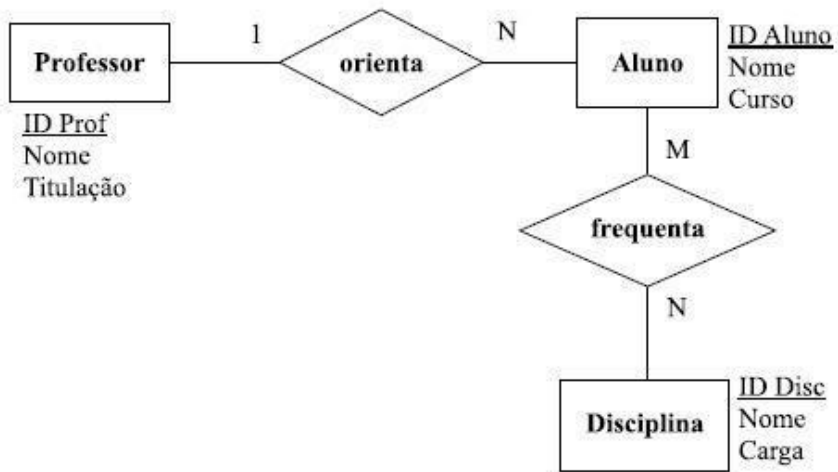
- Domínio de um atributo
 - Descrição de possíveis valores permitidos para um atributo.
 - Ex.: Sexo {M, F}
- Tipo de um Atributo
 - Determina a natureza dos valores permitidos para um atributo.
 - Ex.: inteiro, real, string, etc.

Esquema vs Instância

- Esquema de um Banco de Dados é a **especificação da estrutura** do Banco de Dados
- Instância é o **conjunto de ocorrências** dos objetos de dados de um esquema em um dado momento do tempo

Codigo	Nome	Sigla
1	Tecnologia da Informação	TI
2	Recursos Humanos	RH

Relacionamentos



- As entidades são conectadas umas às outras através de relacionamentos.
- Exemplos:
 - As pessoas moram em apartamentos
 - Os apartamentos formam condomínios
 - Os condomínios localizam-se em ruas ou avenidas
 - As avenidas e ruas estão em uma cidade

Relacionamentos

Relacionamento UM PARA UM (1:1):

- Onde uma entidade X se associa unicamente a uma ocorrência da entidade Y.

Relacionamento UM PARA MUITOS (1:N):

- Onde uma entidade X se associa a várias ocorrências da entidade Y, porém, a entidade Y pode apenas se associar a uma ocorrência da entidade X.

Relacionamento MUITOS PARA MUITOS (N:N):

- Onde a entidade X o pode se associar a várias ocorrências da entidade Y e a entidade Y pode também se associar a várias ocorrências da entidade X.

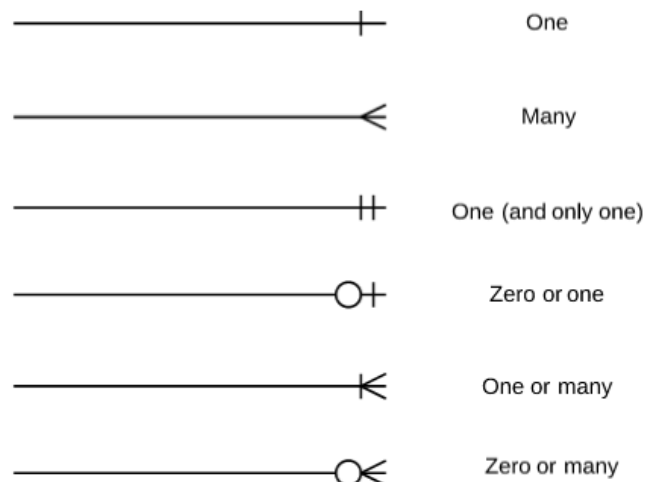
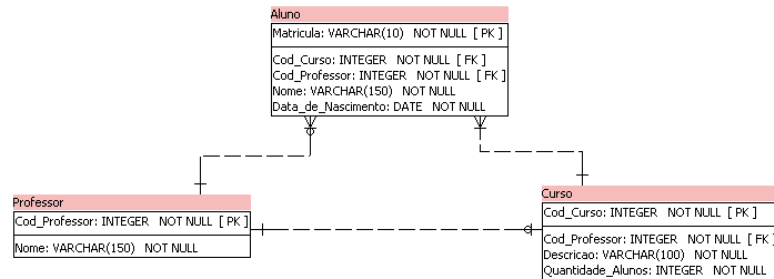
Aluno
Matricula: VARCHAR(10) NOT NULL [PK]
Nome: VARCHAR(150) NOT NULL
Data_de_Nascimento: DATE NOT NULL

Professor
Cod_Professor: INTEGER NOT NULL [PK]
Nome: VARCHAR(150) NOT NULL

Curso
Cod_Curso: INTEGER NOT NULL [PK]
Descricao: VARCHAR(100) NOT NULL
Quantidade_Alunos: INTEGER NOT NULL

Relacionamentos

COMO FICARIAM OS RELACIONAMENTOS DAS ENTIDADES ABAIXO?



Relacionamentos

- Relacionamento entre Aluno e Curso
 - Um curso pode ter vários alunos inscritos e um aluno pode estar inscrito em um curso.
- Relacionamento entre Professor e Curso
 - Um curso pode ser coordenado por um professor e um professor pode coordenar um curso.
- Relacionamento entre Professor e Aluno
 - Um professor pode orientar vários alunos e um aluno pode ser orientado por um professor.



Resumindo!

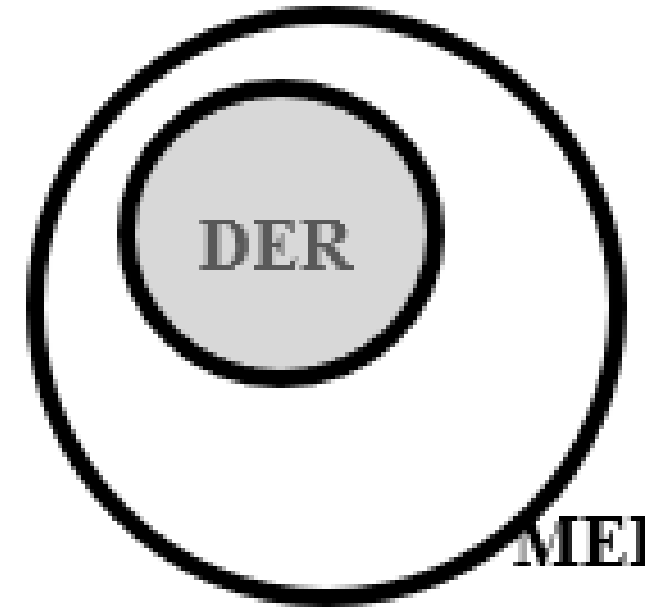
ACESSE:

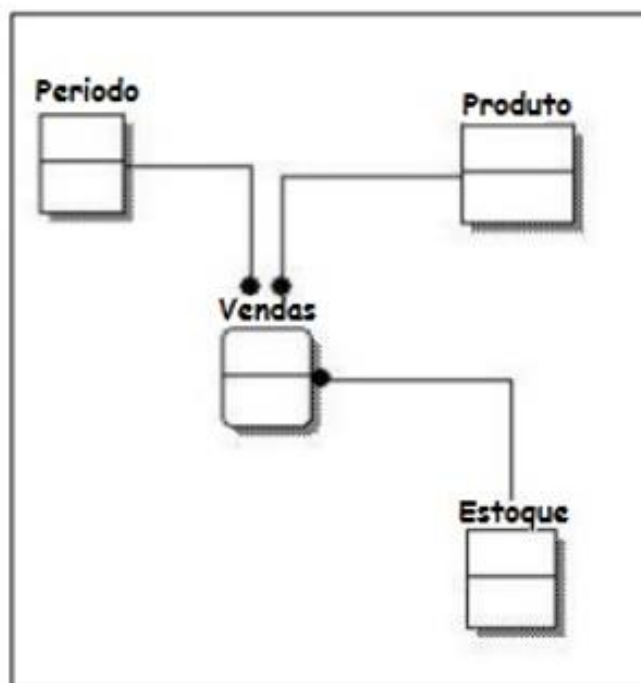
<https://www.lucidchart.com/pages/pt/simbolos-de-diagramas-entidade-relacionamento>

Resumindo mais: MER é diferente de DER!

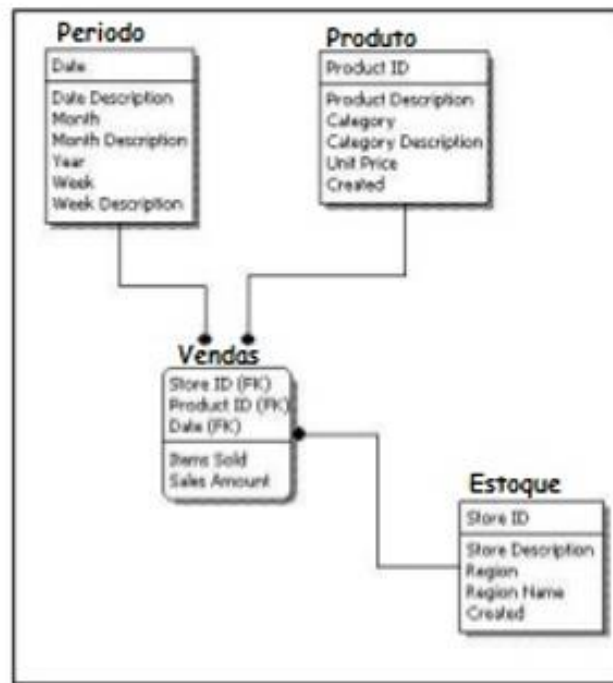
Método de Peter Chen

- Etapas:
 - 1) Análise ambiente - Entidades/Relacionamentos/Atributos
 - 2) Diagrama E-R (DER)
 - 3) Mapeamento E-R (Transforma DER em tabelas)
 - 4) Estrutura dos Registros (define campos)
- Foi concebido para representar a semântica que os dados possuem no mundo real.
- Sua representação gráfica é facilitada através do diagrama entidade-relacionamento (DER)

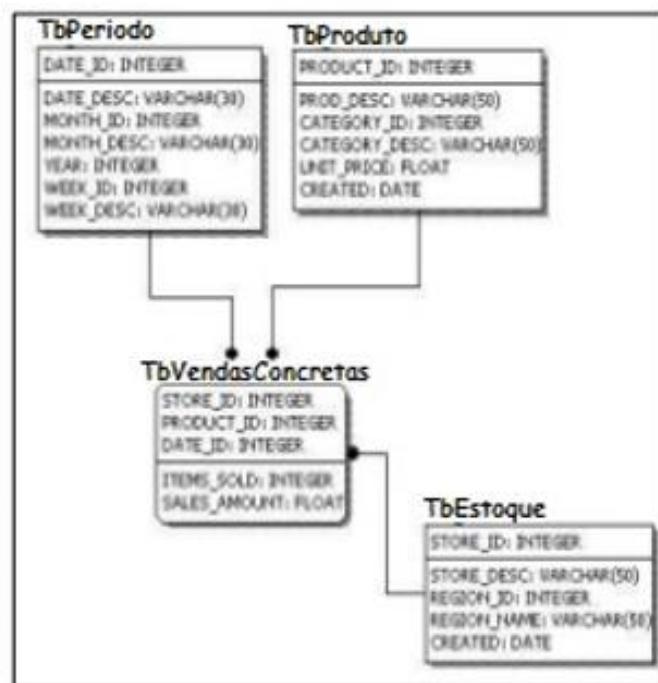




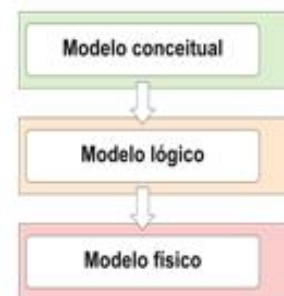
Modelo Conceitual



Modelo Lógico



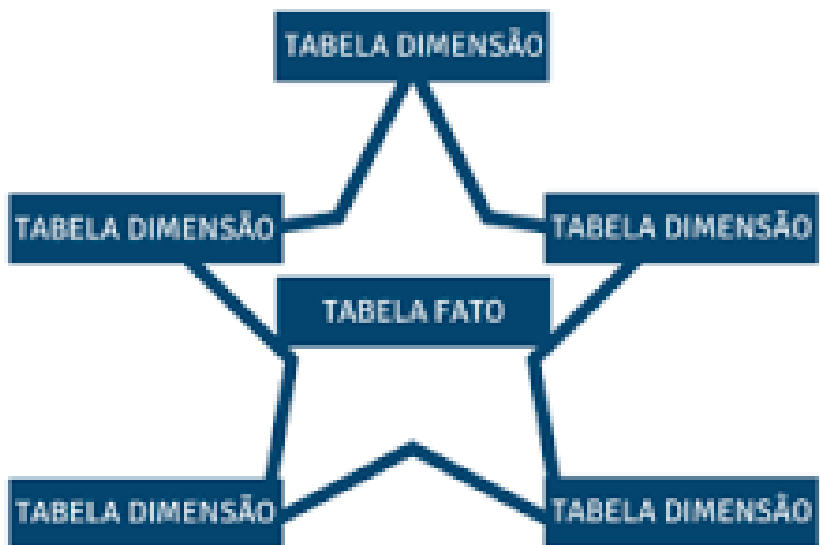
Modelo Físico



Tipos de modelagem então...



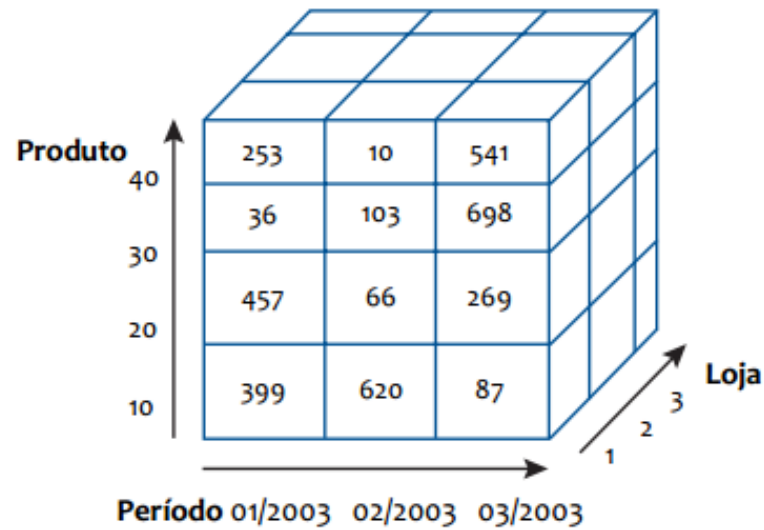
Modelagem dimensional



Modelagem dimensional

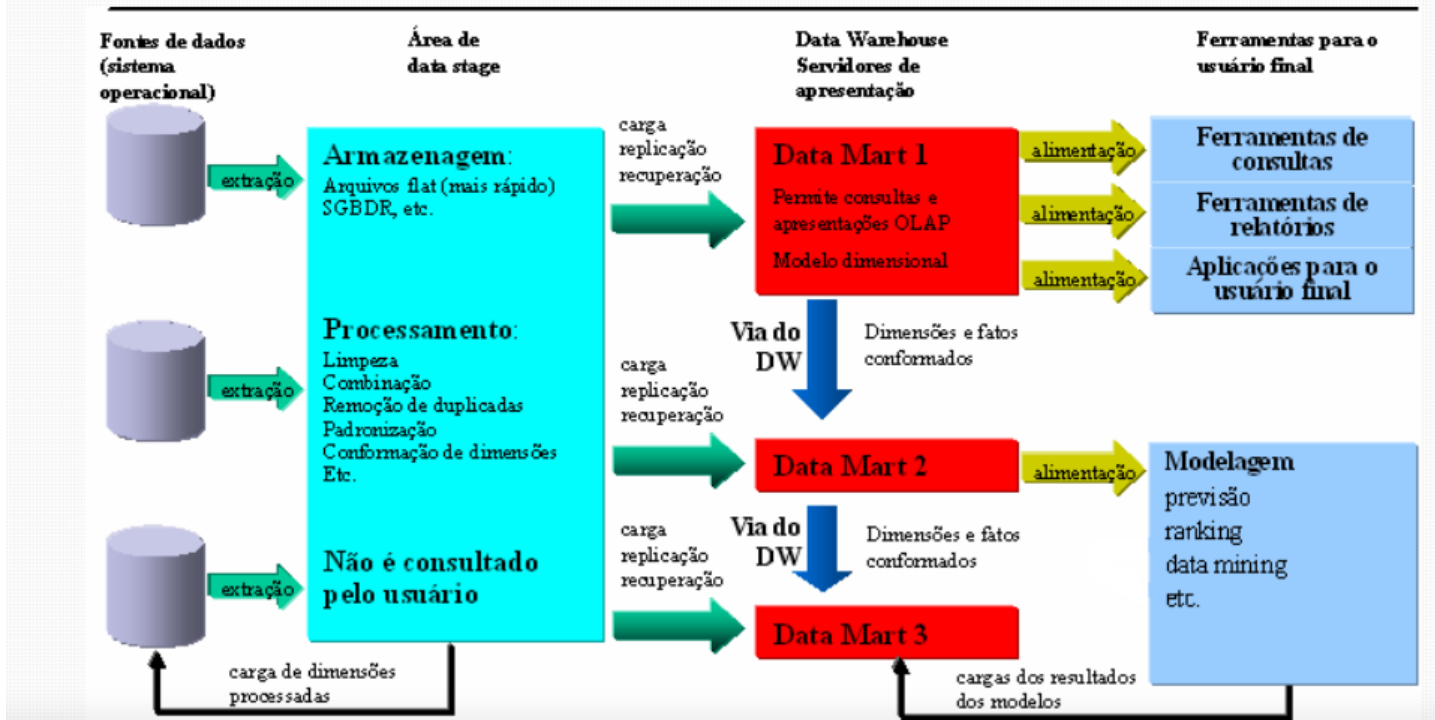
- Outro modelo de modelagem de dados é a Modelagem Dimensional.
- É utilizado na modelagem de dados OLAP (Online Analytical Processing) e tem como objetivo entregar o dado e informação.
- Essa forma de modelagem não está preocupada com a redundância e consistência dos dados, os dados estão desnormalizados, há muita redundância, o banco de dados aumenta de tamanho, mas a obtenção da informação é rápida.
- O foco é a consulta.

Modelagem dimensional

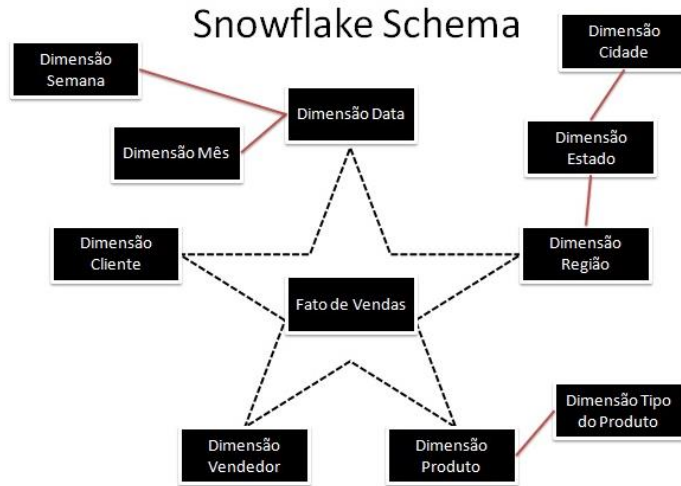


- Data Warehouse (DW), segundo Date (2014), “é um depósito de dados orientado por assunto, integrado, não volátil, variável com o tempo, para apoiar as decisões gerenciais”.
- Um data warehouse é utilizado para armazenar informações relativas às atividades de uma organização de forma consolidada.
- É um Banco de dados organizado para dar suporte à tomada de decisões estratégicas da empresa.
- Enquanto o Data Warehouse usa dados de toda a corporação, os chamados Data Marts (DM) têm objetivo idêntico, mas em geral tratam apenas um assunto ou processo de negócio.

Abaixo podemos visualizar os elementos básicos que compõem as **arquiteturas de um Data Warehouse**.

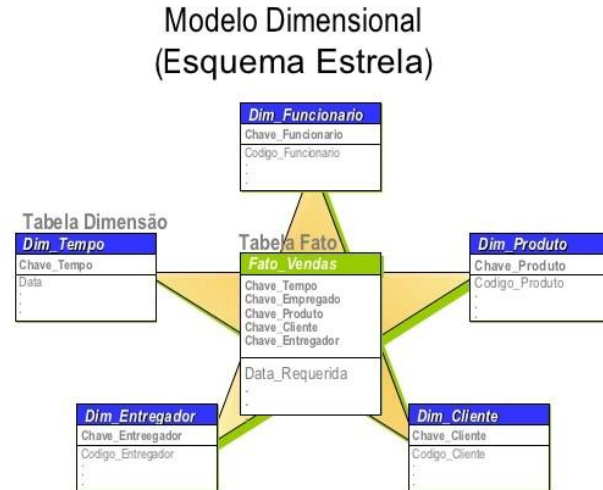


Modelagem dimensional



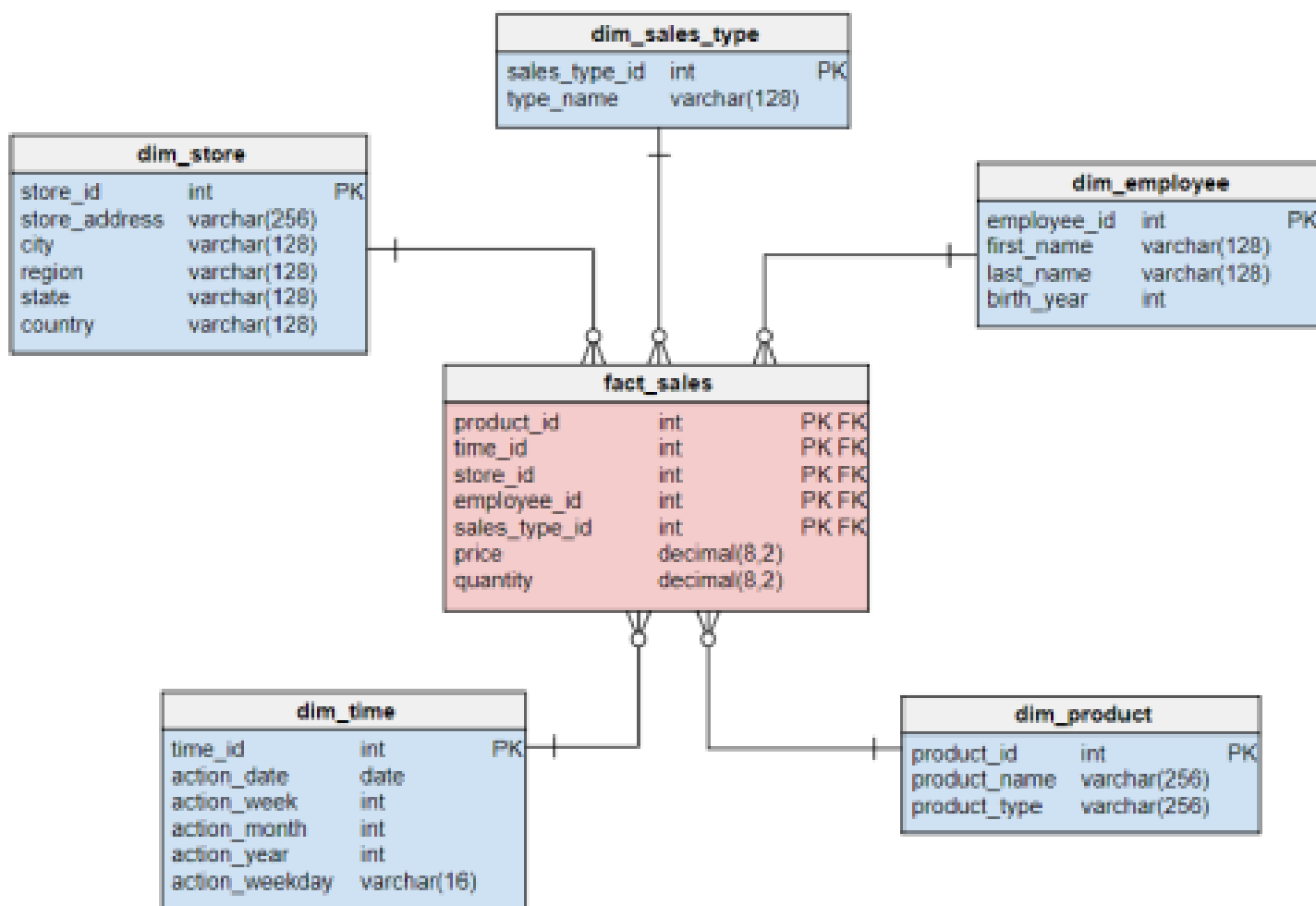
Modelagem dimensional

- A modelagem dimensional é uma das técnicas e conhecimentos mais importantes que você precisa ter para modelar o Data Warehouse, o Data Mart ou o que for.
- Até para utilizar ferramentas de visualização de dados, por exemplo o Power BI da Microsoft, você irá precisar entender de modelagem dimensional.
- Existem dois tipos de metodologias de modelagem de dados dimensional, a Snowflake e a Star Schema, que é a mais utilizada.



Modelagem dimensional

Quando fala-se em cubo, não dá para afirmar, sem conhecer o modelo, se é uma modelagem do tipo estrela ou flocos de neve.



Modelagem dimensional

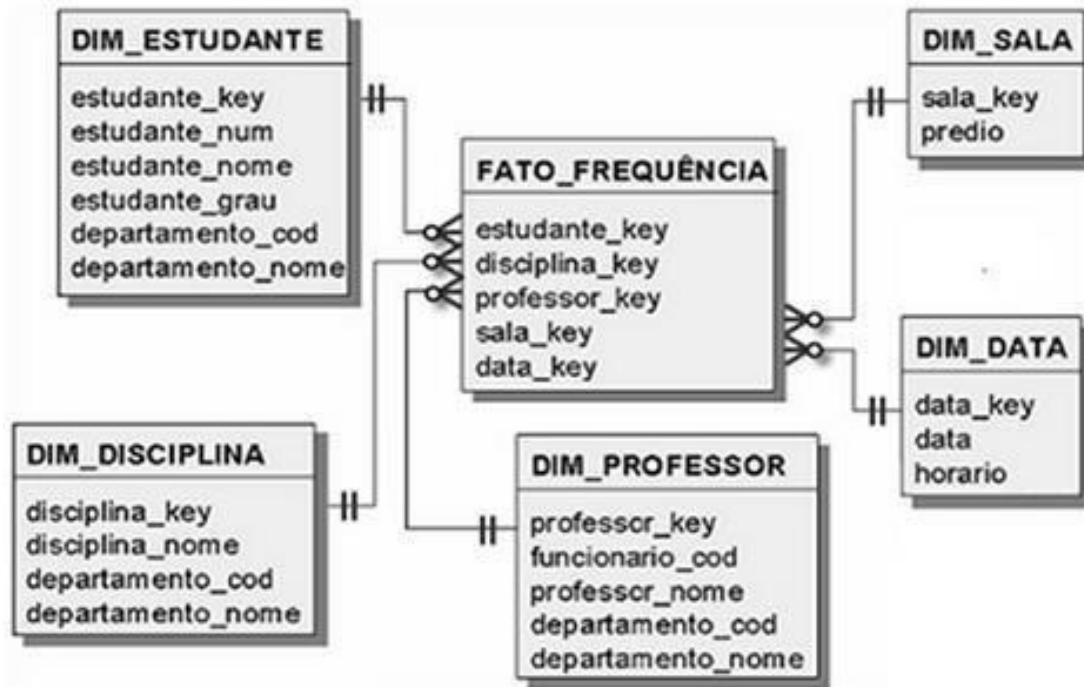


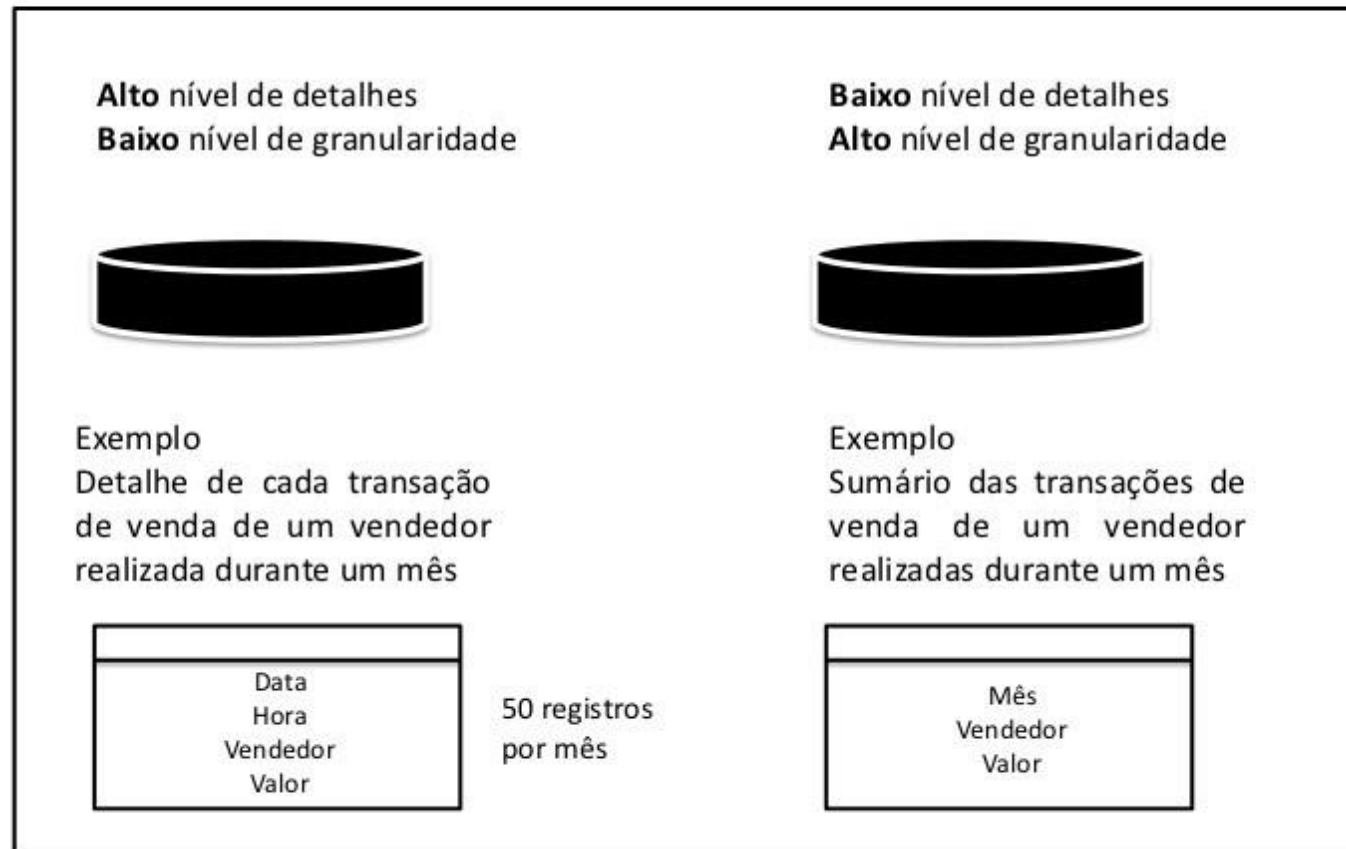
Figura 6 – Modelagem multidimensional.

- O conceito de Star Schema, ou modelo estrela, foi idealizado por Ralph Kimball.
- O Star Schema, na sua essência, é composto no centro por uma tabela Fato, que é rodeada por Dimensões.
- Por isso tem o nome de Star Schema, porque parece uma estrela.
- As entidades no modelo Dimensional são as tabelas Fatos e Dimensões.

Modelagem dimensional

- As dimensões contém as características de um fato
 - Por exemplo, quando um cliente fez a compra, onde a compra foi feita, que produto foi comprado, ou quem era o cliente? Já a tabela Fato armazena o que ocorreu, é o fato propriamente dito.
- A tabela Fato está sempre ligada a duas ou mais dimensões
 - Não existe tabela Fato com menos de duas dimensões
- A tabela Fato conecta-se com as dimensões e isso forma o modelo dimensional
- Ela também é a principal tabela no modelo dimensional

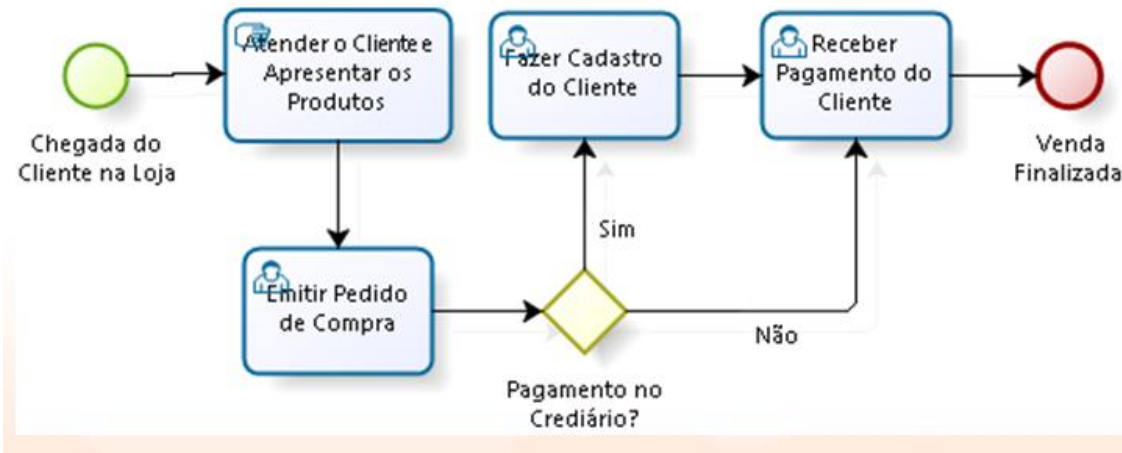
Granularidade dos Dados



Modelagem dimensional

- O processo para construção da Modelagem Dimensional é composto de 4 passos, são eles:
- 1. Identificar o Processo de Negócio
- 2. Identificar a Granularidade
- 3. Identificar as Dimensões
- 4. Identificar os Fatos

Modelagem dimensional



1. Identificar o Processo de Negócio

- Cada empresa tem Processos de Negócios específicos definidos em suas Estratégias de Negócios
- Os sistemas transacionais são responsáveis por suportar os processos de negócios
- Uma ou mais tabelas Fatos são utilizadas para modelar cada Processo de Negócio



Modelagem dimensional

2. Identificar a Granularidade

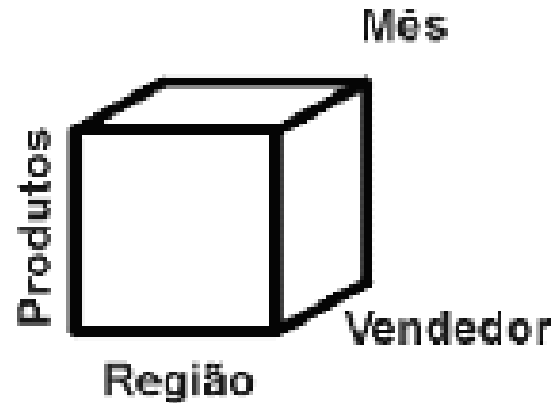
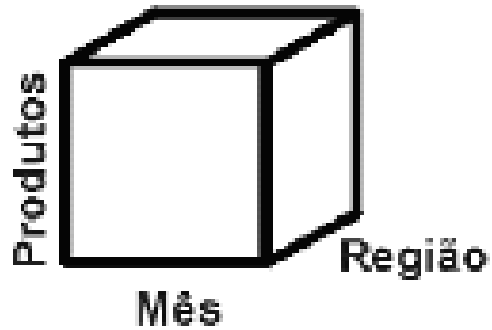
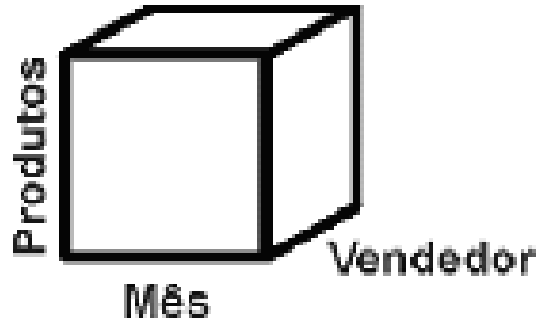
- Nível de detalhe, ou seja, o que uma linha da tabela Fato representa.
- Por exemplo, para o processo Vendas, serão agrupadas as vendas por mês, por semana, por dia, por hora, ou por minuto?
- As recomendações são para começar pelo nível mais baixo da informação, lembrando sempre que o menor nível é o que tem a maior ocupação de espaço e não pode complicar nenhuma das análises.
- O nível da informação deve ser o necessário e não para atender futuras expectativas.

Granularidade de Dados

Granularidade – O que considerar?



Modelagem dimensional



3. Identificar as Dimensões

- Dimensões são as descrições que a informação pode ter, essas podem ser filtradas, agrupadas, selecionadas
- A granularidade define o conjunto de dimensões que serão utilizados
- Mais adiante teremos uma sessão específica falando de dimensão

Modelagem dimensional

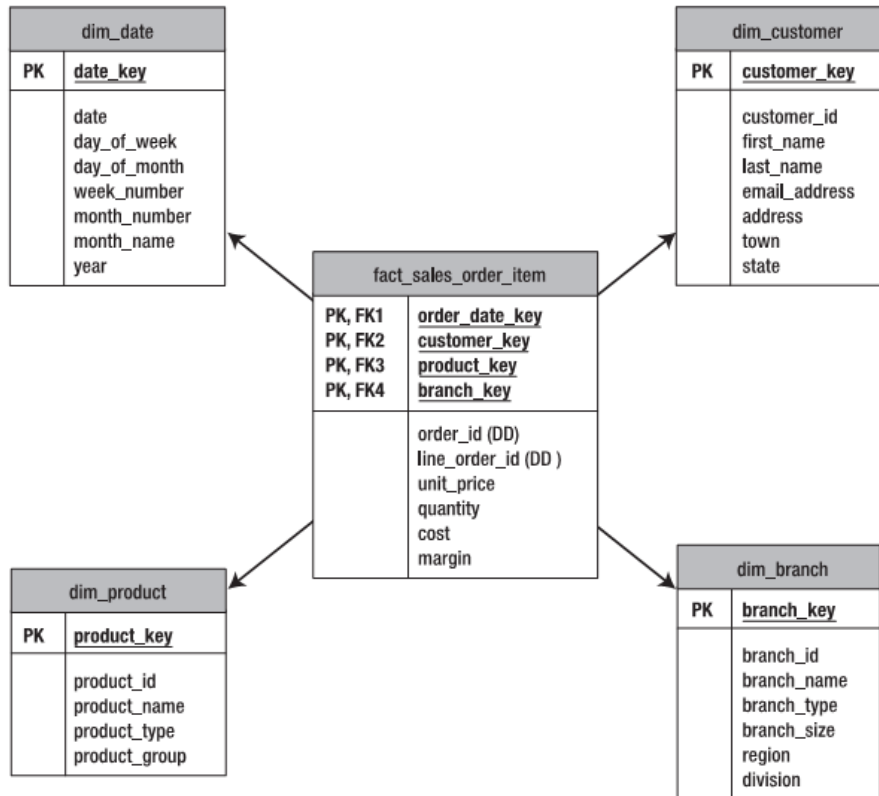


Figure 1-4. Star schema dimensional data store

4. Identificar os Fatos

- São as métricas resultantes do Processo de Negócio, quantidades, valores vendidos, etc.
- As tabelas Fatos também podem ser derivadas ou calculadas.
- Mais adiante teremos uma sessão específica falando de Fatos.

Modelagem dimensional



É muito importante no processo de modelagem Dimensional considerar os dados disponíveis.

É comum, durante a modelagem dimensional, constatar que os sistemas transacionais da organização não contêm todos os dados necessários para a modelar o processo de negócio definido.

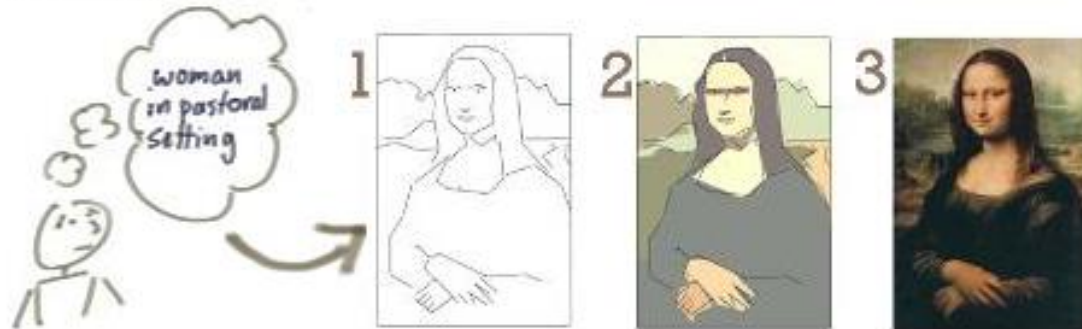
Nesses casos os sistemas transacionais devem ser adequados, ou devem ser obtidos dados externos para que as análises possam ser feitas da maneira desejada.

Modelagem dimensional

Incremental

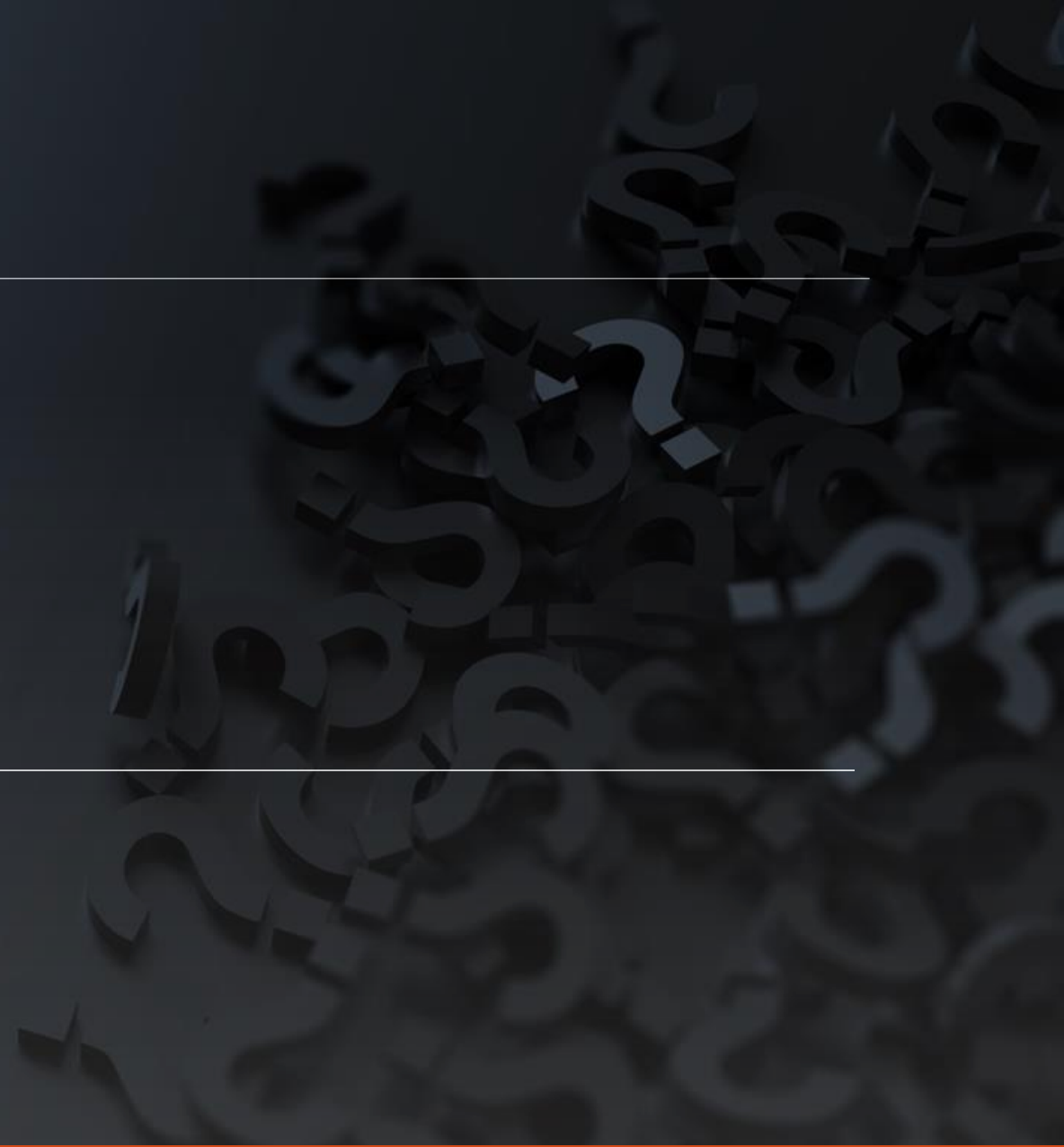


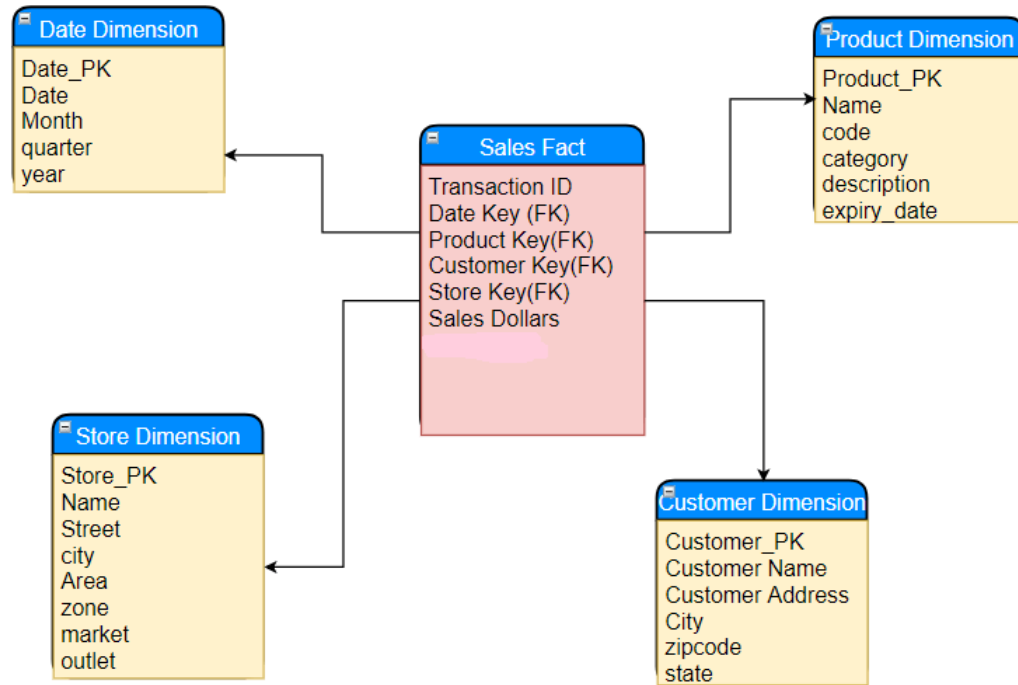
Iterative



- O processo de construção de um modelo de dados é um processo INCREMENTAL e ITERATIVO que, geralmente, não é finalizado em um único passo
- Para se chegar a um modelo completo, muitos passos e transformações ocorrem com base no modelo inicial
- Visite o modelo de dados de sua solução com frequência!
- Vídeo: [\(530\) Iterativo Incremental - YouTube](#)

Quiz!

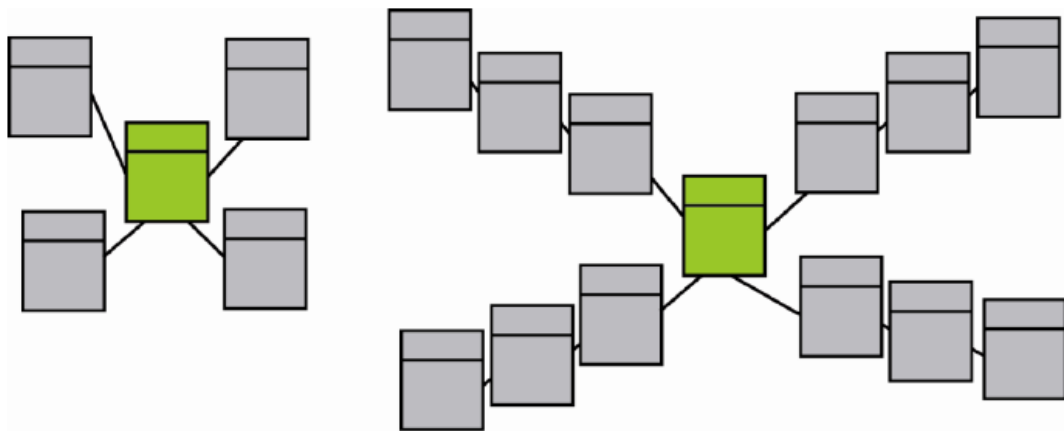




Additive Facts

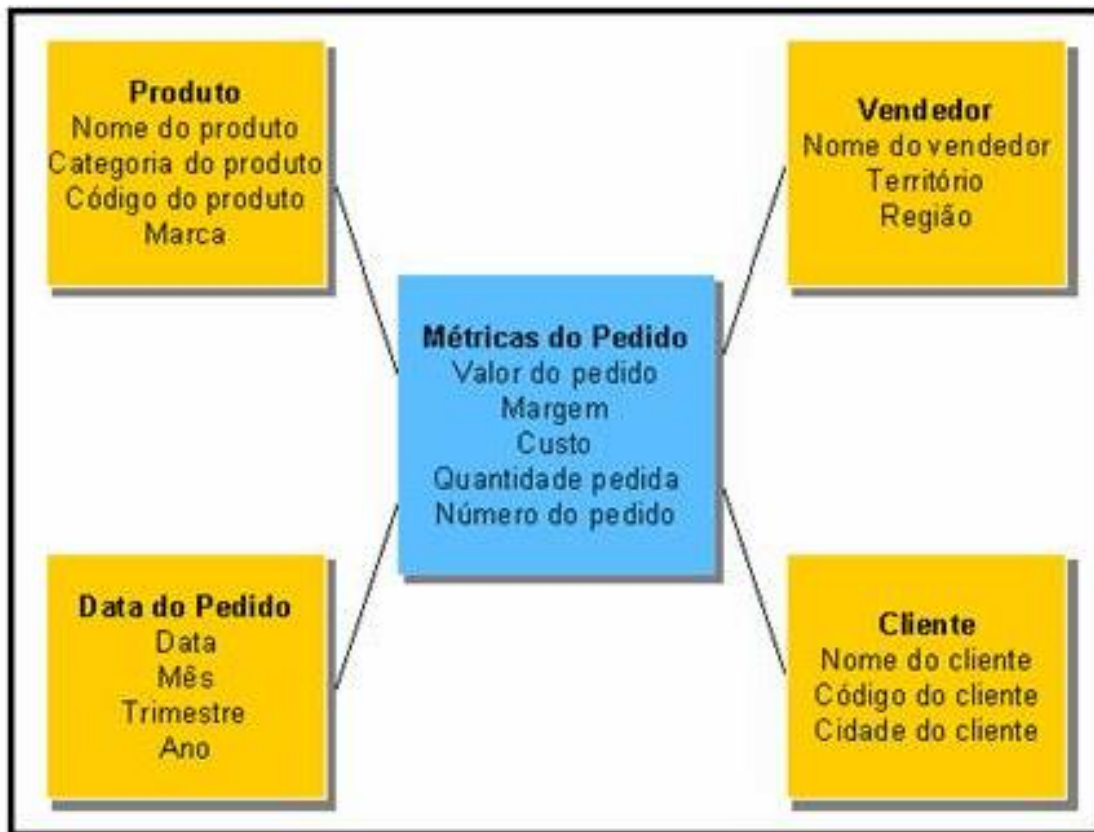
Fatos

- A Fato é a tabela da modelagem Dimensional que armazena os dados quantitativos no DW.
- Nelas são modeladas as métricas e KPIs que serão cruzadas com as Dimensões, concebendo, assim, informações significativas para a análise do usuário.
- A Fato armazena as medições necessárias para avaliar o assunto pretendido, ou processo de negócio.
- O conteúdo histórico no DW, contendo longo período de tempo, fica depositado na Fato.



Fatos

- O relacionamento, em geral, entre Fatos e Dimensão é de um para muitos.
- A imagem ao lado permite uma visualização simples da junção Fato e Dimensão na modelagem de dados dimensional:



Fatos

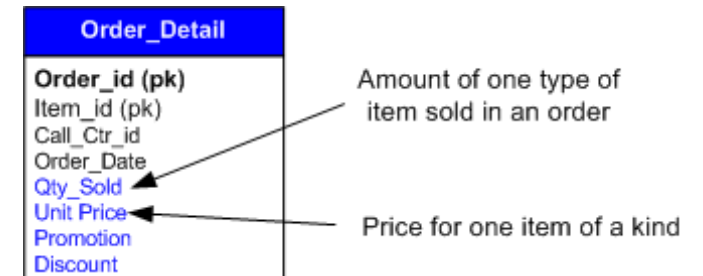
- Exemplo de modelagem dimensional com o processo de negócio Vendas
- Ao modelar a(s) tabela(s) de Fatos, deve-se ter em mente os seguintes pontos:
 - a tabela Fato deve ter ao menos uma dimensão data associada a ela, às vezes, múltiplas dimensões datas estarão ligadas em uma única tabela Fato.
 - A chave primária é composta pela junção da chave de cada dimensão.

Fatos

- Existem dois tipos de granularidades fundamentais que classificam as tabelas Fato: Cumulativa e Snapshot, as outras são consideradas apenas variações das mesmas.
- A Cumulativa descreve o que aconteceu em um período de tempo.
 - Um exemplo é o total de vendas por produto por loja por mês.
 - Esse tipo de tabela Fato contém em sua maioria métricas aditivas.
- A Snapshot descreve o estado das coisas em um instante particular do tempo e geralmente é composta de métricas semiaditivas e não-aditivas.
 - Pode-se considerar como exemplo saldo de estoque, saldos de conta corrente, poupança e investimentos por cliente por mês.

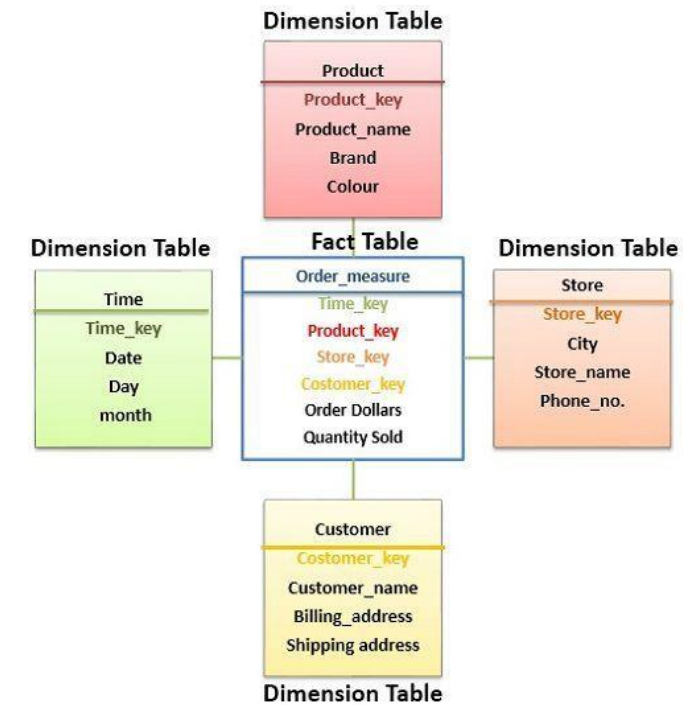
Fatos

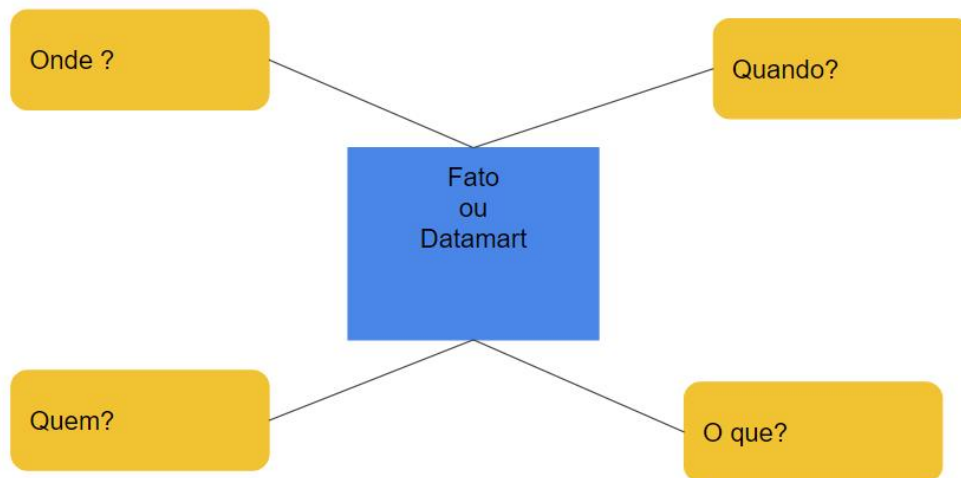
- Independente de sua granularidade, cada métrica em uma tabela Fato deve estar exatamente no mesmo nível de detalhe.
- Quando você mistura fatos representando muitos níveis de granularidade em uma mesma tabela Fato, você estará criando confusão para os usuários de negócios e tornando as aplicações de BI vulneráveis a erros de valores ou outros resultados incorretos.



Dimensões

- As Dimensões armazenam os dados qualitativos de um DW, são os descritores dos dados da Fato.
- É a Dimensão que permite a visualização das informações por diversos aspectos e perspectivas.
- As tabelas Dimensão contêm vários atributos que descrevem em detalhes todas as características que possam definir e serem úteis para futuras tomadas de decisão.

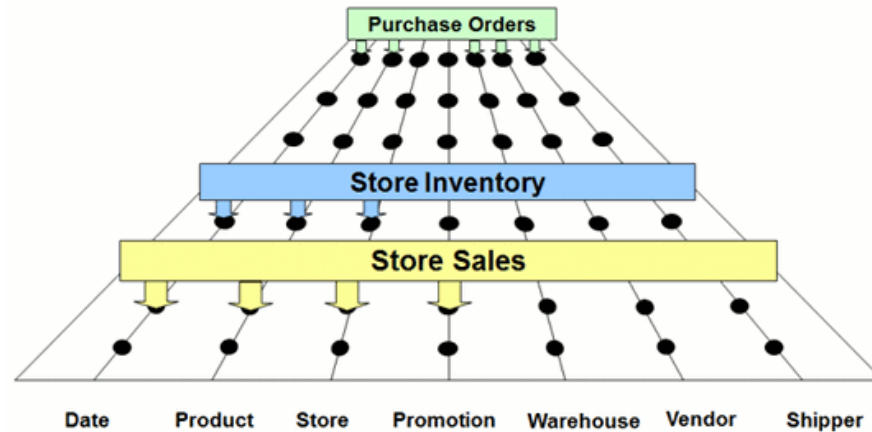




Dimensões

- Ao modelar as dimensões de uma Fato, tenha em mente 4 perguntas guia:
 - Quando o fato ocorreu?
 - Onde ocorreu?
 - Com quem ocorreu?
 - E o que estava envolvido?
- Com essas 4 perguntas é possível mapear a maiorias das dimensões de análise de uma Fato para a tomada de decisão.
- Lembre-se, não existe fato sem a dimensão tempo.
- A figura mostra um esquema gráfico para as perguntas.

Dimensões e Processos de Negócios



- Ao lado uma imagem de barramento de dimensão, que mapeia todas as dimensões contidas num DW e mostra como essas estão relacionadas com o processos de negócios e fatos.

BUSINESS PROCESSES	COMMON DIMENSIONS						
	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X

Test Drive de Modelagem

Na pasta da disciplina de Modelagem, no Github.

- Procurar o arquivos: Exercício 1 – Modelagem Dimensional.pdf

Na próxima aula...

Discutiremos os componentes da modelagem dimensional.

- Métricas, Hierarquias e Atributos

