SÃO PAULO
GOVERNO DO ESTADO
SÃO PAULO SÃO TODOS

DESENVOLVIMENTO DE SOFTWARE MULTIPLATAFORMA

Disciplina: IBD-016 – BANCO DE DADOS - NÃO RELACIONAL

Aula 02: Introdução aos conceitos de Data Warehouse

Data 15/08/2024

Prof. Me. Anderson Silva Vanin

Data Warehouse

- Definição
- Ambiente
  - Ferramentas ETL
  - Data Marts
- Arquiteturas
  - Global
  - Data Marts independentes
  - Data Marts integrados
- Implementação
  - Top Down
  - Botton up
  - Combinada
- Metadados

Introdução

- A informação é o melhor recurso do qual empresas podem dispor para tomar decisões
- Obtida analisando históricos sobre vendas, clientes, produtos, etc
- Dados conflitantes de fontes diversas podem gerar informações desencontradas

Introdução

- A quantidade de dados a serem considerados cresce com a expansão do negócio e com o passar do tempo...

- Data Warehouses auxiliam a resolver esses problemas ao prover montantes gigantescos de dados temporais integrados para posterior análise!

Introdução

- Criado pela IBM na década de 60 com o nome Information WareHouse
- Relançado diversas vezes sem sucesso
- O nome Data WareHouse foi dado por William Inmon, considerado o pai desta tecnologia
    - Tornou-se viável com o surgimento de novas tecnologias para armazenar e processar uma grande quantidade de dados.

Definição

Conjunto de dados **agrupados por assunto**, **integrados**, variável em relação ao **tempo** e **não volátil**, que serve de suporte para o processo de **tomada de decisões.**

Definição

- **Orientado a Assunto**
  - Um Data Warehouse está sempre orientado ao redor do principal assunto da organização
  - Ao contrário de aplicações clássicas, orientadas por processos/funções
- **Integrado**
  - Os dados criados dentro de um ambiente de Data Warehouse são integrados
  - A integração beneficia com a convenção consistente de nomes, estrutura consistente de códigos etc

Definição

- **Não volátil**
  - Os dados nunca são excluídos nem alterados de um Data Warehouse
- **Variante no tempo**
  - Data Warehouse apresenta os dados com seu posicionamento em relação ao tempo

Comparativo com BD operacional

Aspecto	BD Operacional	Data Warehouse
Usuários	Funcionários	Alta administração
Utilização	Tarefas Cotidianas	Decisões estratégicas
Padrão de Uso	Previsíveis	Difícil de prever
Princípio de Func.	Com base em transações	Com base em análise de dados
Valores de dados	Valores atuais e voláteis	Valores históricos e imutáveis
Detalhamento	Alto	Sumarizado
Organização dos dados	Orientado a aplicações	Orientado a assunto

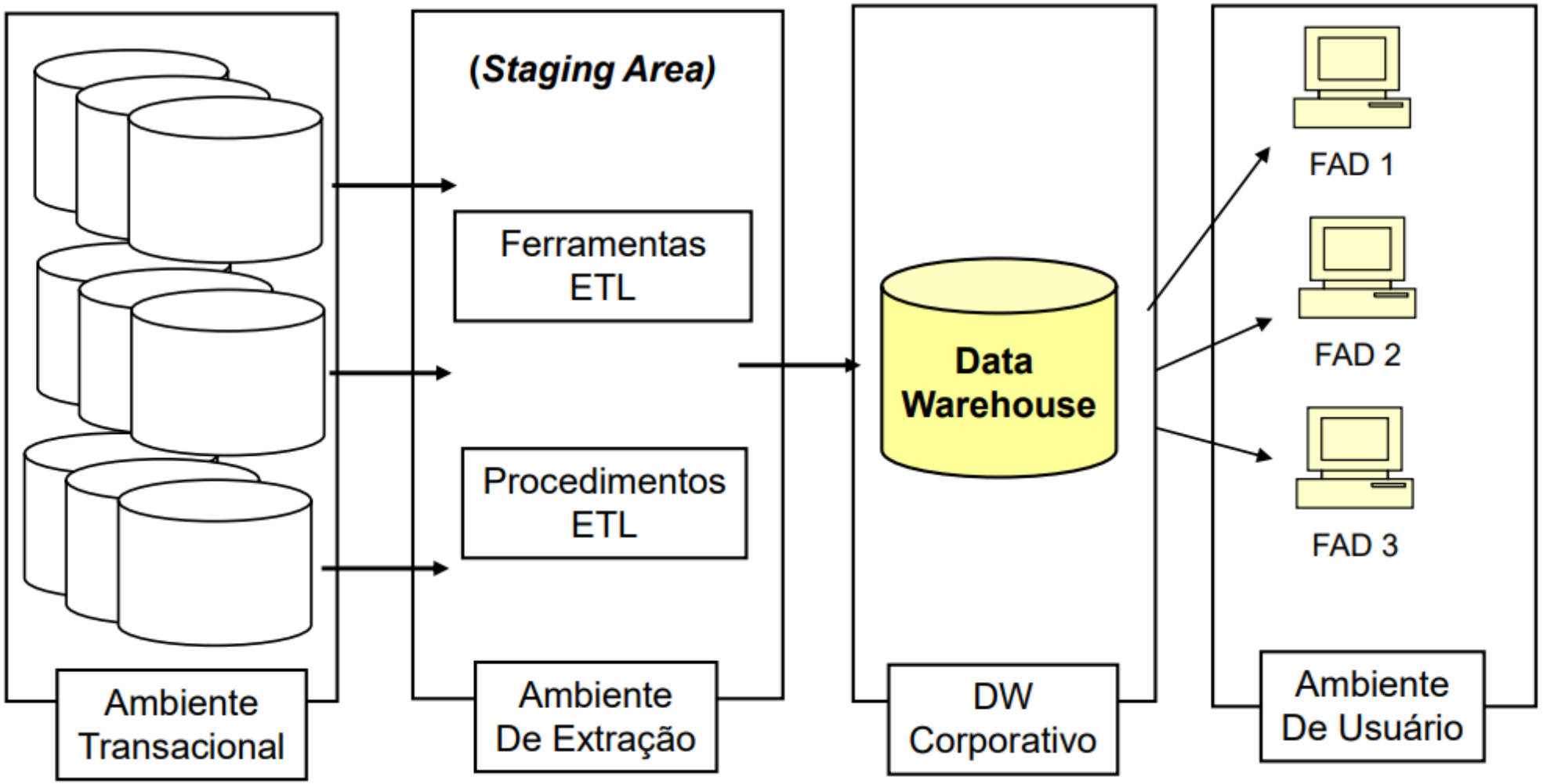
Comparativo com BD operacional

Algumas diferenças adicionais do Data Warehouse para um BD operacional

- Permitem a redundância de dados
- Buscas complexas e ad hoc (personalizadas pelo usuário)
- Modelagem de dados multidimensional

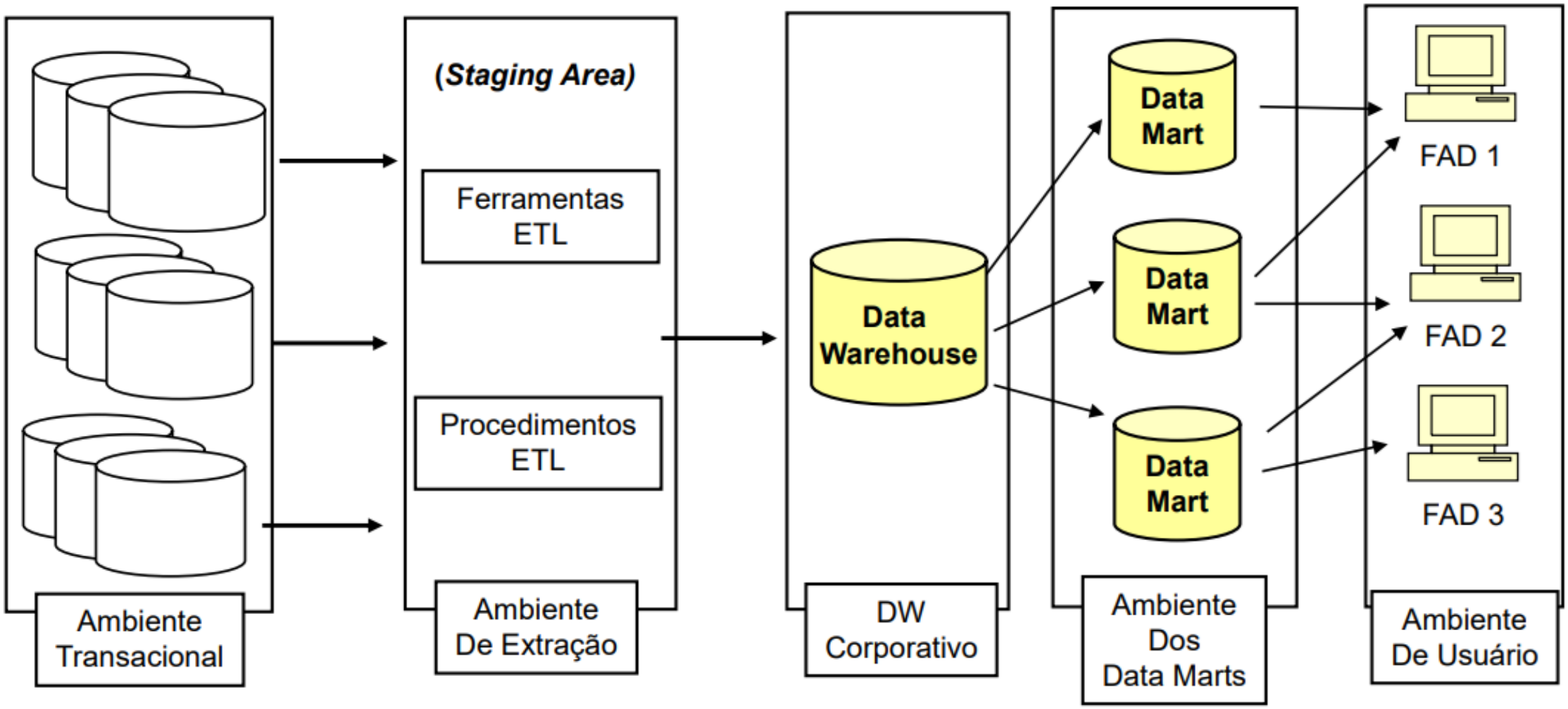
Ambiente de Data Warehouse

- Centralizado



Ambiente de Data Warehouse

- Com Data Marts



Ambiente de Data Warehouse

Extraction, Transformation and Load

- Consiste da integração e limpeza dos dados
- Integração: consolidação dos dados de diversas origens
- Limpeza: rejeição de valores inválidos

Ambiente de Data Warehouse

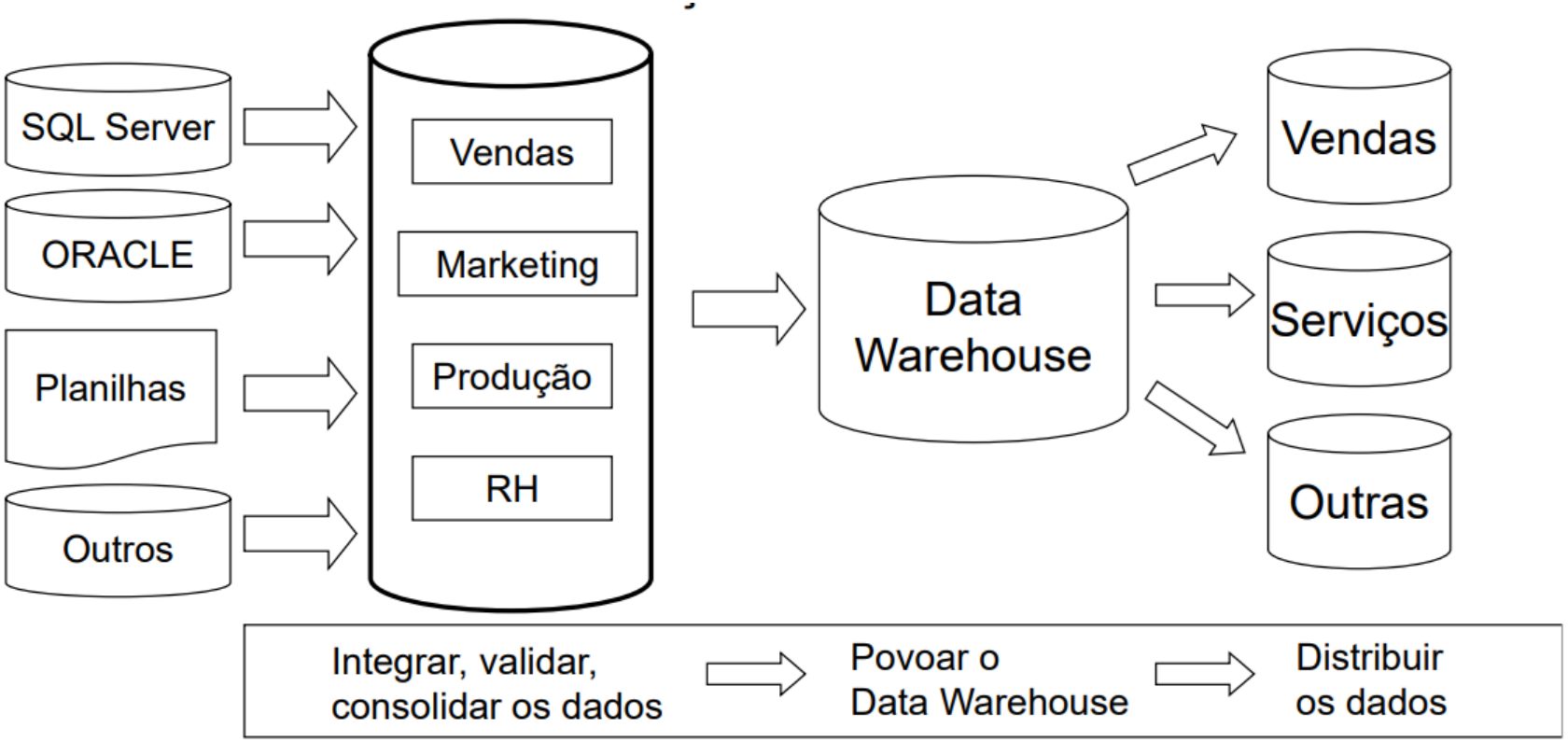
Extraction, Transformation and Load

- Os processos ETL consomem 70% do tempo de desenvolvimento em um projeto de DW
- Estes processos são específicos para cada organização
- Opcionalmente, pode-se ter uma segunda área intermediária, chamada Operational Data Store (ODS)

Ambiente de Data Warehouse

Extraction, Transformation and Load

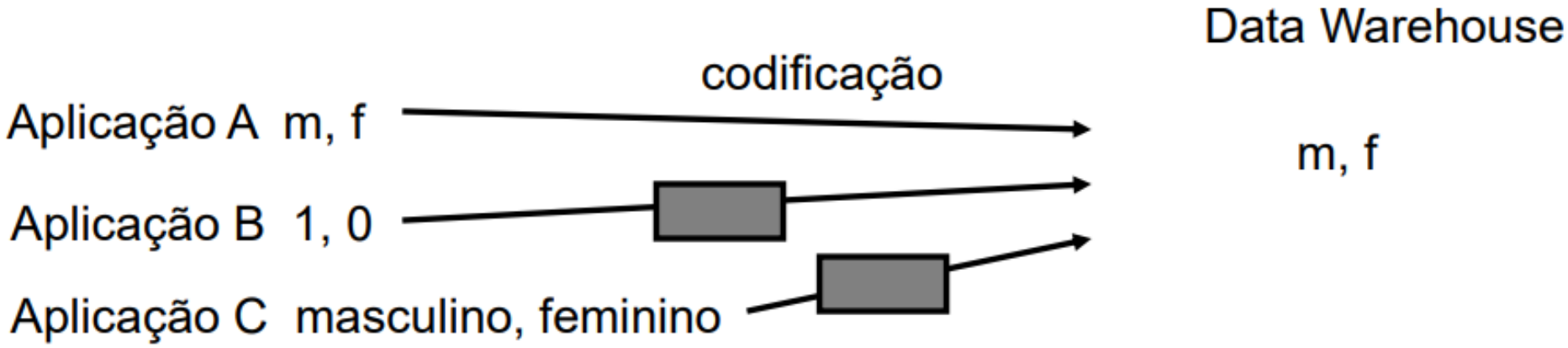
- Carga – receber os dados de diversos Sistemas de Processamento de Transações



Ambiente de Data Warehouse

Extraction, Transformation and Load

- Transformação e integração - processo de formatação e modificação de dados extraídos de várias origens para transformá-los em informações úteis ao Data Warehouse
- Os dados de origem são consistentes mas apresentados de diferentes formas



Ambiente de Data Warehouse

Extraction, Transformation and Load

- Transformação e integração

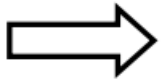
mudança

Nome_Cliente	Categoria	Compras
Barr, Adam	II	17,60
Chai, Sean	IV	52,80
Melia, Erin	VI	8,82
...	...	...

Nome_Cliente	Categoria	Compras
Barr, Adam	2	17,60
Chai, Sean	4	52,80
Melia, Erin	6	8,82
...	...	...

Primeiro	Ultimo	Categ	Compras
Adam	Barr	2	17,60
Sean	Chai	4	52,80
Erin	Melia	6	8,82
...		...	...

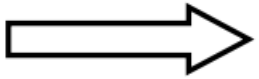
junção



Nome_Cliente	Categ	Compras
Barr, Adam	2	17,60
Chai, Sean	4	52,80
Melia, Erin	6	8,82
...	...	...

Nome_Cliente	Preço	Qtde
Barr, Adam	0,55	32
Chai, Sean	1,10	48
Melia, Erin	0,98	9

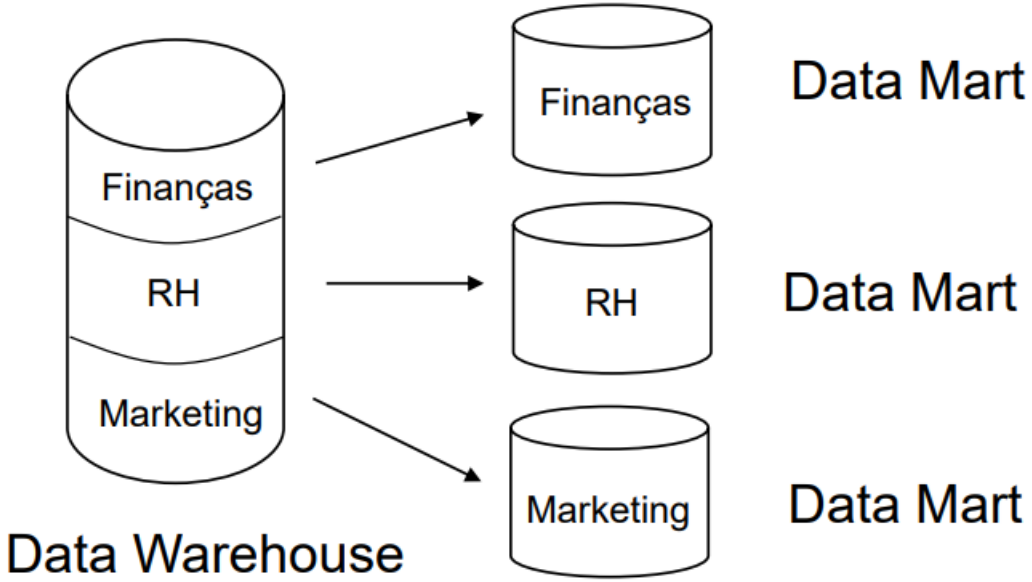
cálculo



Nome_Cliente	Preço	Qtde	Total
Barr, Adam	0,55	32	17,60
Chai, Sean	1,10	48	52,80
Melia, Erin	0,98	9	8,82
...	...	...	...

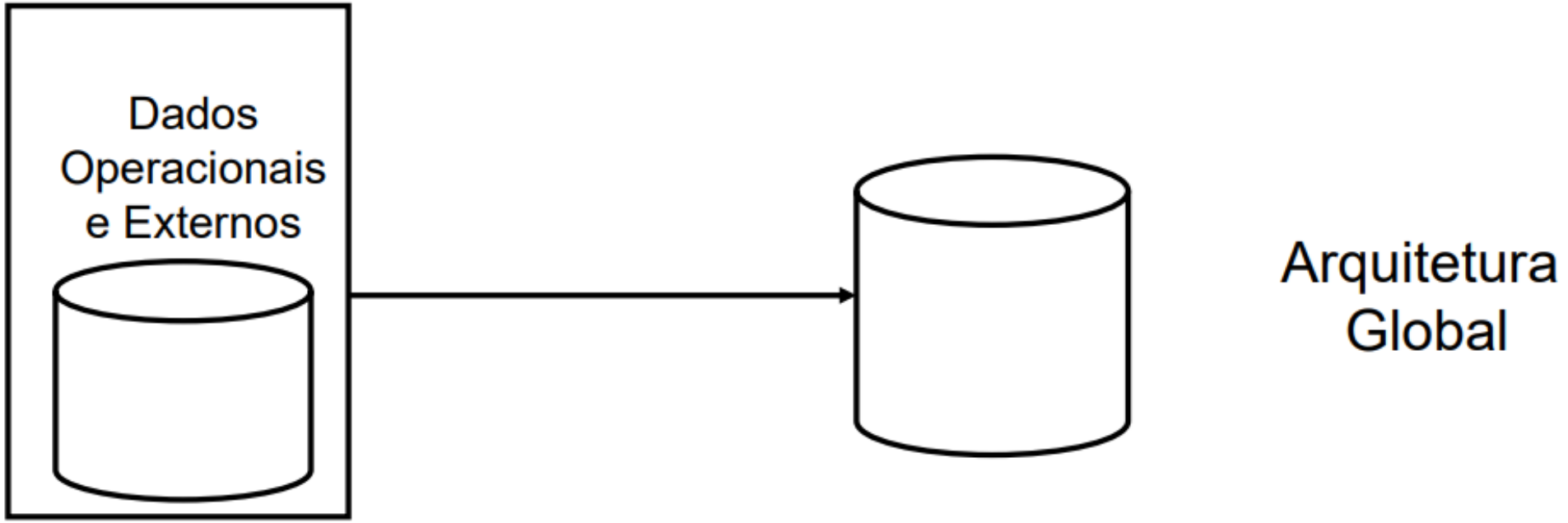
Ambiente de Data Warehouse – Data Mart

- Data Mart - Subconjunto lógico de um Data Warehouse, um Data Warehouse setorial
  - Geralmente descritos como um subconjunto dos dados contidos em um Data Warehouse extraído para um ambiente separado



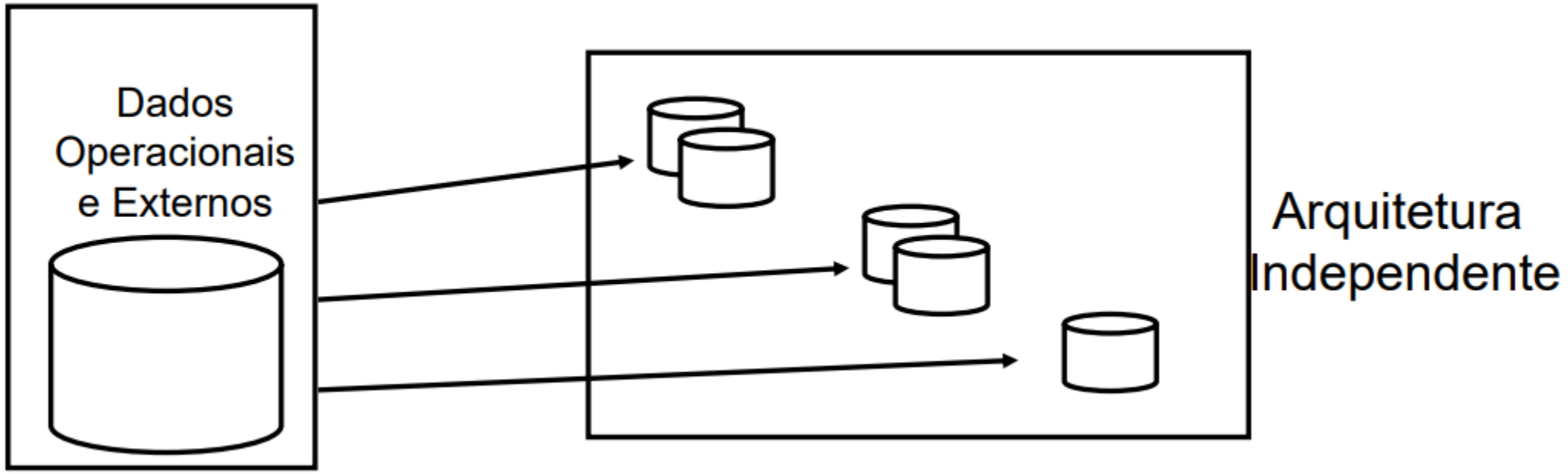
Arquitetura de Data Warehouses

- **Arquitetura global**
  - Utiliza um repositório comum de dados, integrado, utilizado por toda a organização



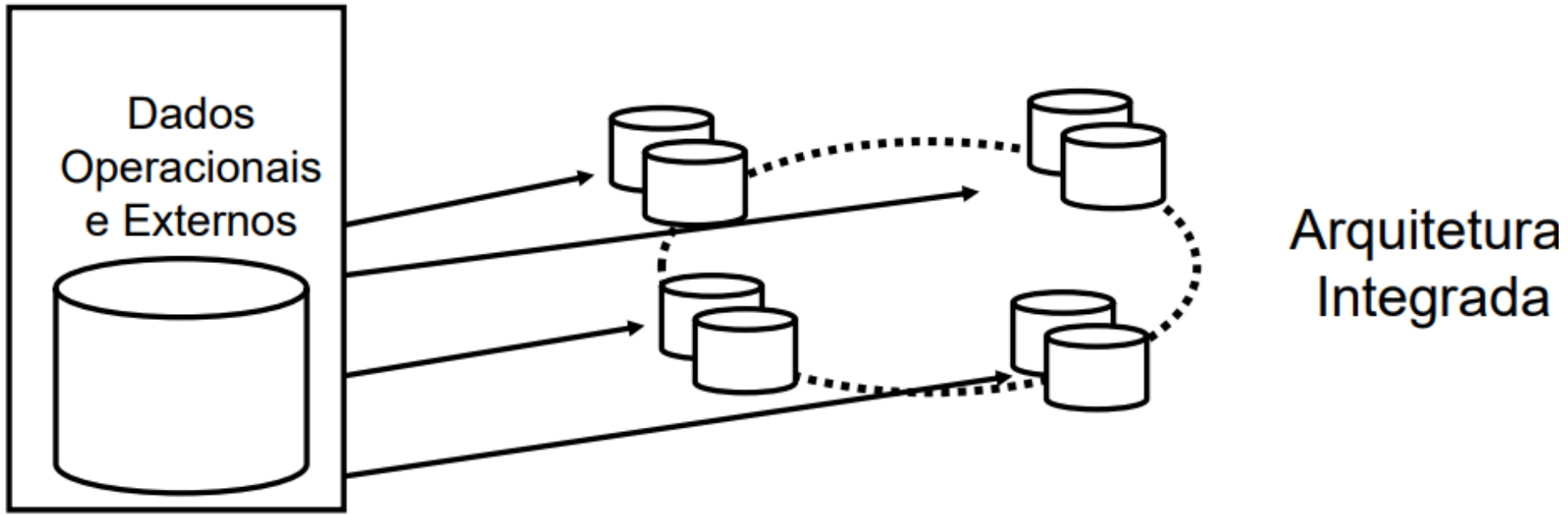
Arquitetura de Data Warehouses

- **Arquitetura de data marts independentes**
  - Possui um data mart para atender a cada departamento em específico
  - Não se tem acesso aos data marts de outros departamentos



Arquitetura de Data Warehouses

- **Arquitetura de data marts integrados**
  - Possui um data mart para atender a cada departamento em específico
  - Os dados são compartilhados entre os data marts de diferentes departamentos



Implementação de Data Warehouses

- **Abordagem Top-down**
  - O modo como os dados serão armazenados e consultados nasce do DW e posteriormente são distribuídos entre os Data Marts
  - Tem objetivo de atender às necessidades da organização como um todo e não departamentos isolados
  - Modelo mais comum de implementação
  - Demorada implementação e resultado apenas a longo prazo

Implementação de Data Warehouses

- **Abordagem Bottom-up**
  - Parte dos Data Marts até compor o DW por completo
  - Maior dificuldade na padronização dos dados
  - Implementação mais rápida e manutenção mais fácil devido ao menor tamanho das partes

Implementação de Data Warehouses

- **Abordagem Combinada**
  - Combina características de ambas abordagens
  - Planejamento geral da estruturação do DW para toda a organização (conforme Top-down)
  - Desenvolvimento dos data marts de forma graduada, apresentando funcionalidades parciais
  - A criação de cada data mart é padronizada para facilitar a integração dos dados

Metadados

- **“Dados sobre dados”**
- Possuem papel de grande importância nos DW
  - Especialmente na fase de desenvolvimento, onde especificam os dados de variadas fontes..

Modelagem - Introdução

- Características do Modelo Entidade-Relacionamento
  - Foco em aplicações transacionais
  - Foco no armazenamento momentâneo (não-histórico) da informação
  - Tende a um grande número de tabelas
  - Eficiente apenas para consultas simples e diretas

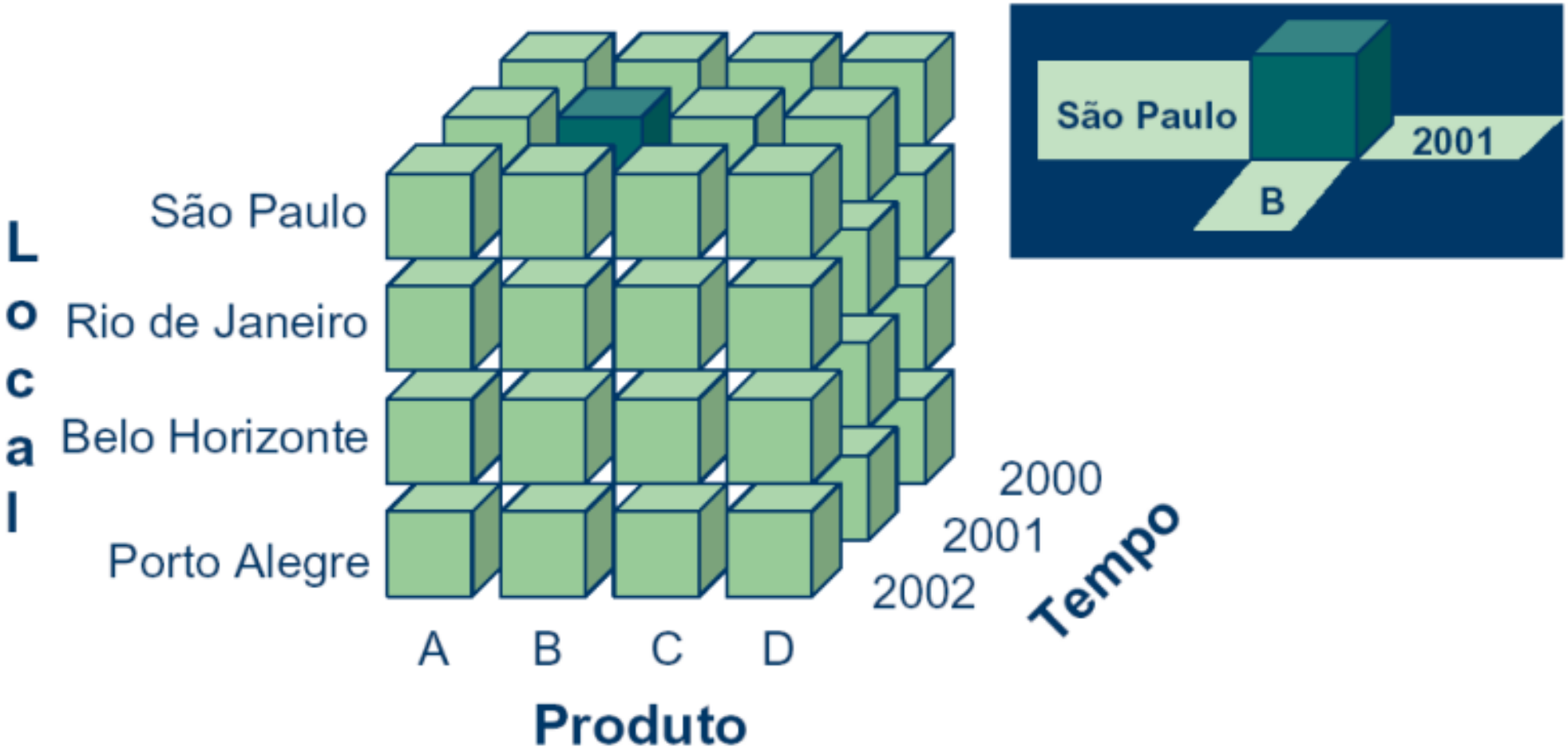
Modelagem - Introdução

- Necessidades em um ambiente de Data Warehouse
  - Foco em aplicações gerenciais
  - Análise histórica das informações
  - Visão ampla das informações (sumarizações, cruzamentos)
  - Visualizar os dados sob diferentes perspectivas (consultas complexas)
- Para implementar um Data Warehouse necessitamos de um novo modelo, diferente do ER tradicional...

Modelagem Multidimensional

- Representar os tipos de dados por uma estrutura chamada cubo de dados
  - Células contêm valores
  - Lados definem as dimensões de análise
- Normalmente também refere-se a cubo de dados mesmo quando há mais de 3 dimensões
  - No entanto, o termo técnico para tal estrutura é **Hipercubo**

Modelagem Multidimensional



Visão Relacional X Visão Multidimensional

- Visão relacional
  - Volume de vendas de uma loja de instrumentos musicais por instrumento e estado

Instrumento	Estado	Qtde. Vendas
Violão	MG	140
Violão	RJ	100
Violão	SP	150
Guitarra	MG	140
Guitarra	RJ	120
Guitarra	SP	80
Bateria	MG	30
Bateria	RJ	20
Bateria	SP	50

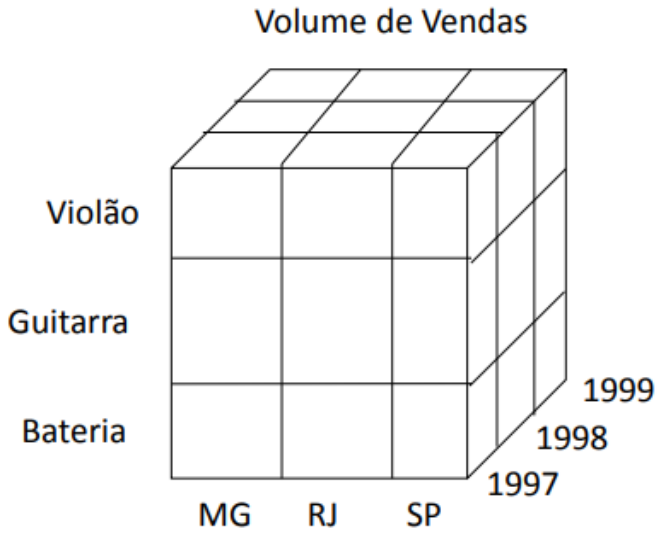
Visão Relacional X Visão Multidimensional

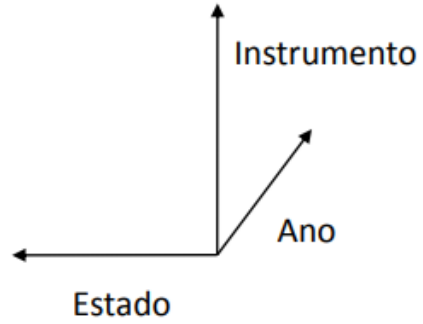
- Visão multidimensional
  - Volume de vendas de uma loja de instrumentos musicais por instrumento e estado

Volume de Vendas

Instrumento	MG	RJ	SP
Violão	140	100	150
Guitarra	140	120	80
Bateria	30	20	50

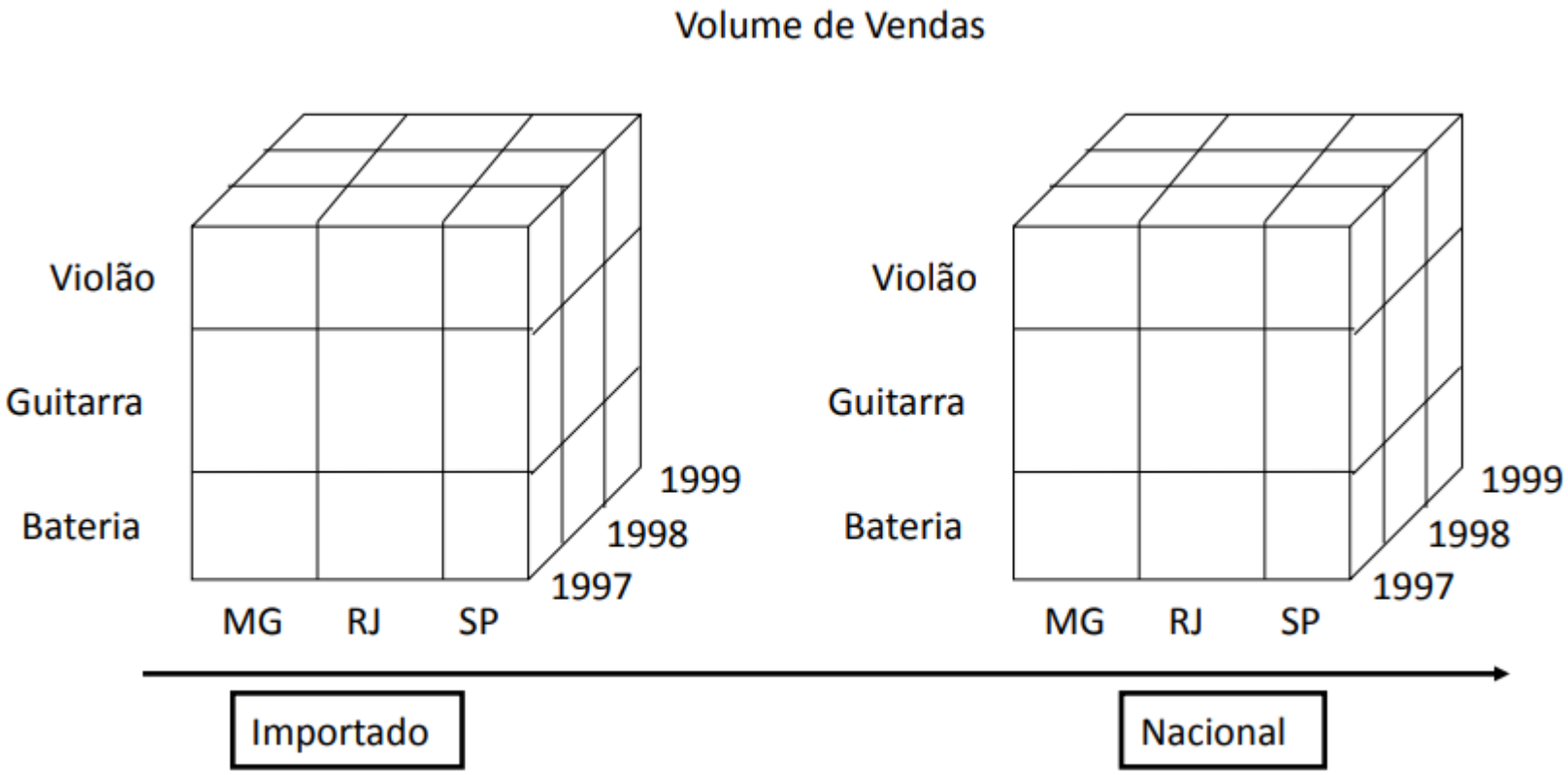
Estado





Visão Relacional X Visão Multidimensional

- Visão multidimensional
  - Hipercubo



Modelagem Multidimensional

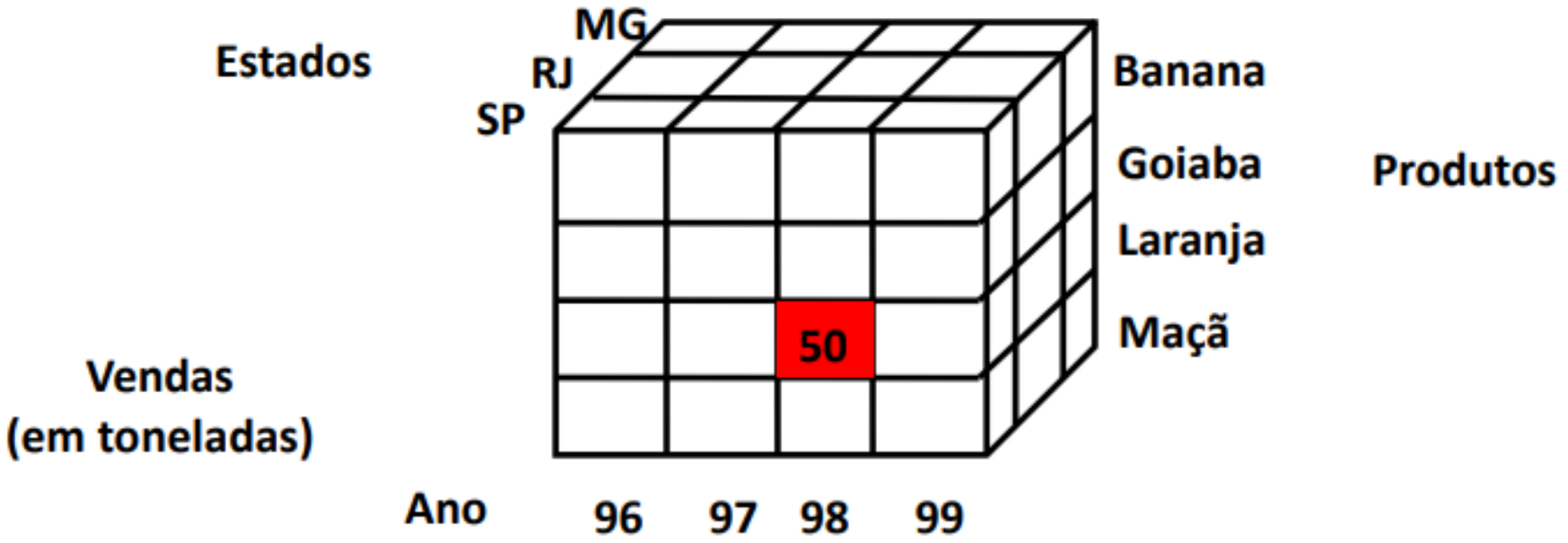
- Elementos básicos
  - Fatos – aquilo que pode ser representado por valores numéricos. Esse conjunto de valores é também chamado métricas ou medidas. Ex.: Vendas
  - Dimensões – determinam o contexto no qual os fatos são analisados. Ex.: Local, Ano e Produto
  - Variáveis – atributos numéricos que representam os fatos. Ex.: Valor (R$) das vendas, Unidades vendidas

Modelagem Multidimensional

- Star Schema (Esquema Estrela)
  - Forma de dispor as tabelas do banco para simular um banco de dados multidimensional
  - Composto por uma tabela dominante, chamada tabela de fatos, rodeada de tabelas auxiliares, chamadas tabelas de dimensão;
  - A tabela de fatos conecta-se às demais por múltiplas junções e as tabelas de dimensões se conectam com apenas uma junção à tabela de fatos.

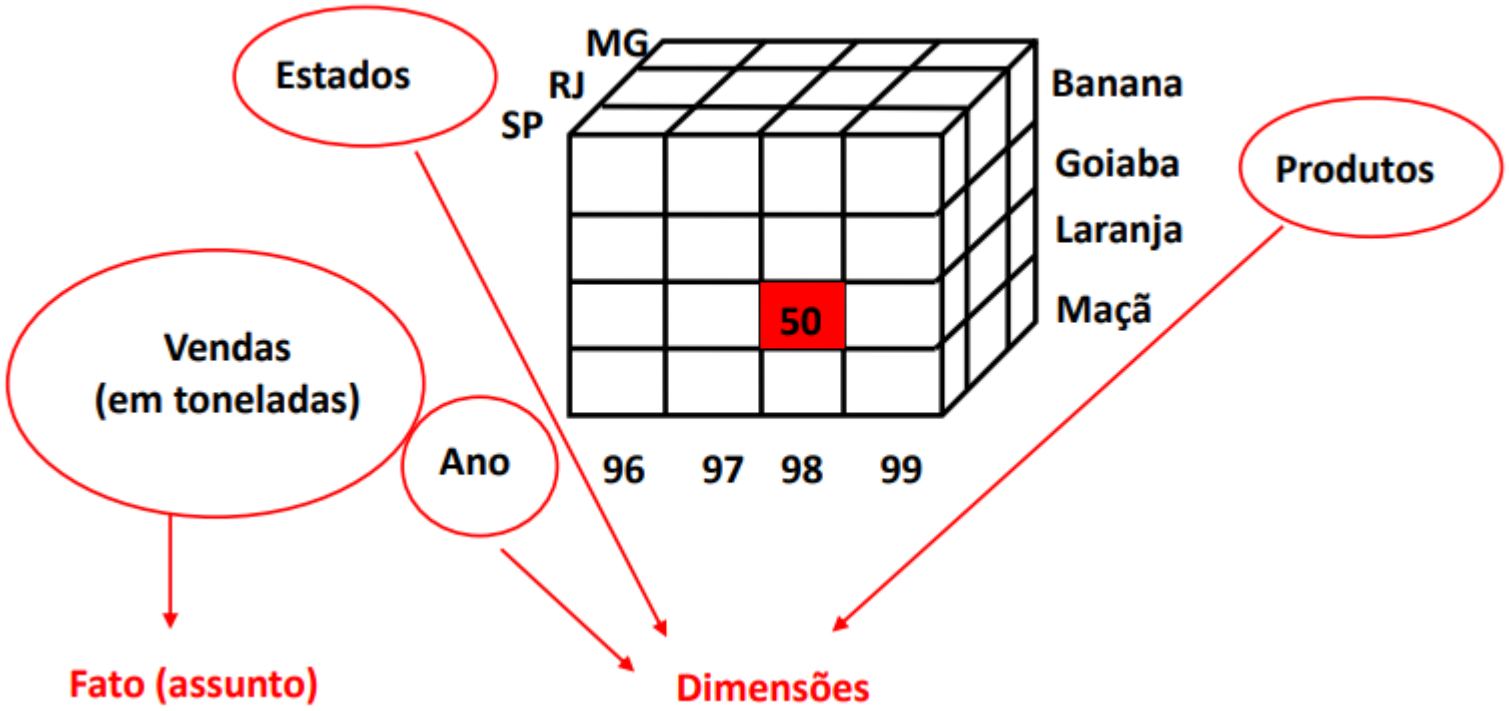
Modelagem Multidimensional

- Star Schema (Esquema Estrela)



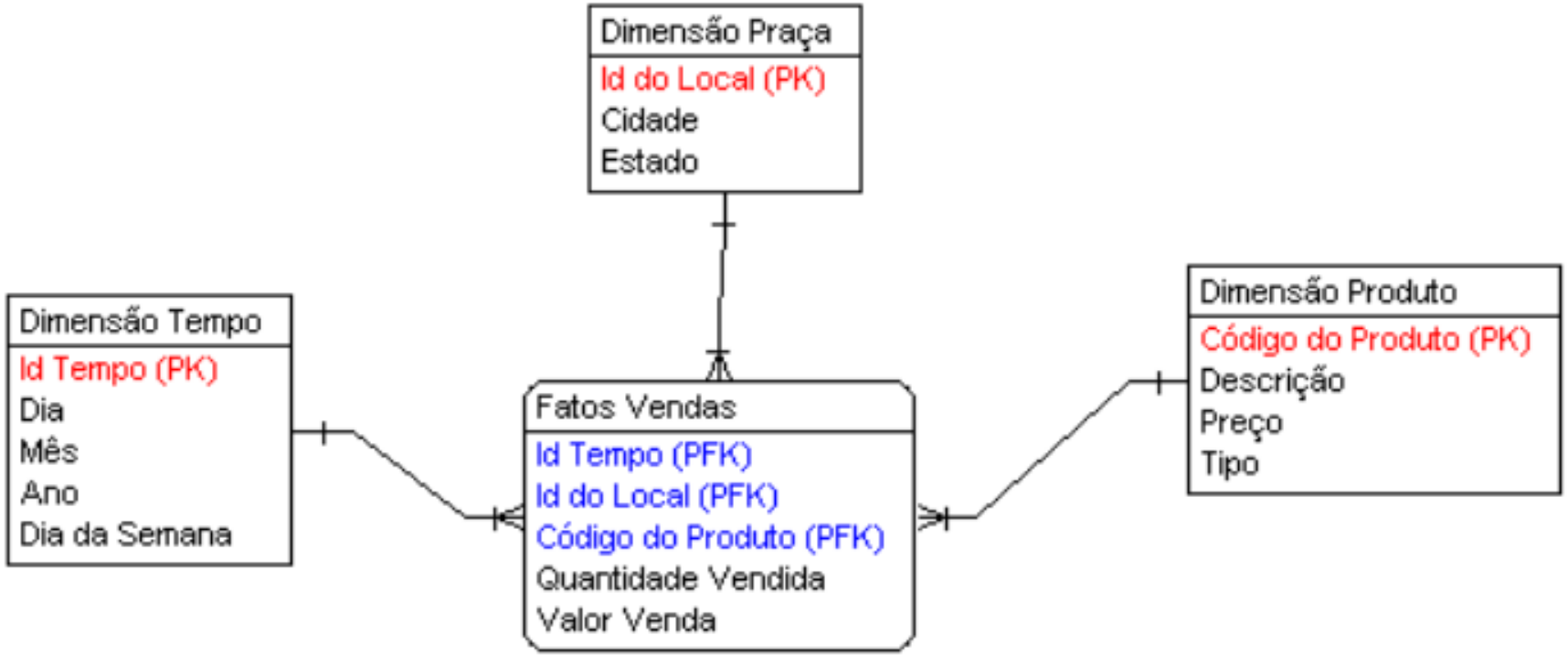
Modelagem Multidimensional

- Star Schema (Esquema Estrela)



Modelagem Multidimensional

- Star Schema (Esquema Estrela)



Modelagem Multidimensional

- Star Schema (Esquema Estrela)
- Exemplo de tabela de dimensão resultante
  - Produto

Id do Produto	Descrição	Preço	Tipo
101	Espaguete	10	Massa
102	Hamburguer	5	Carne
103	Talharim	15	Massa
104	Peito de Frango	20	Carne

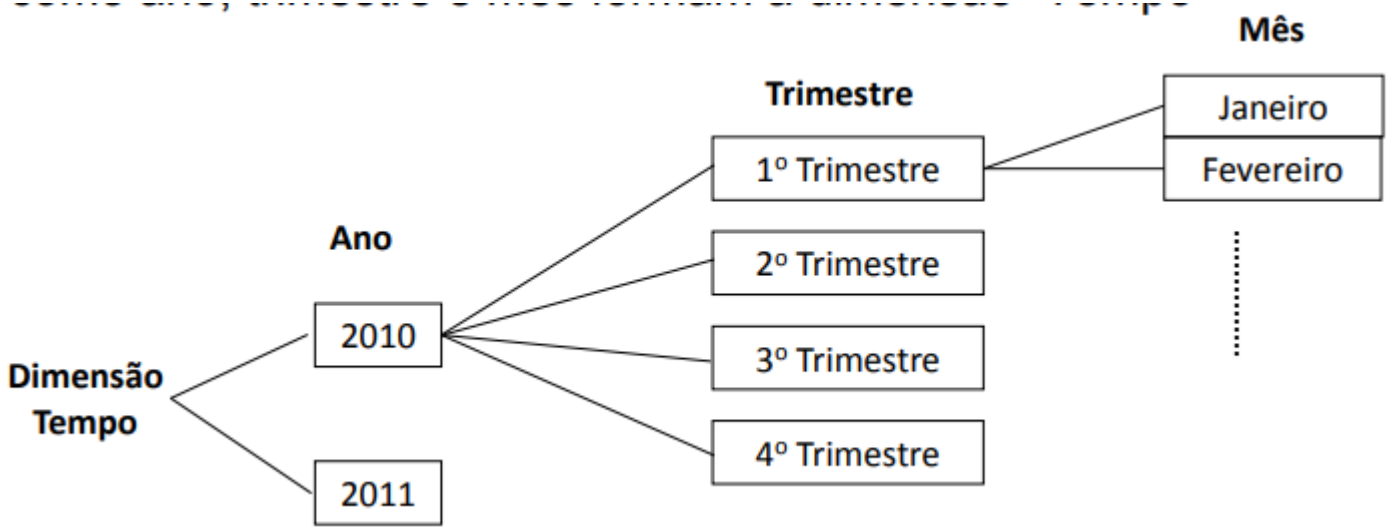
Modelagem Multidimensional

- Star Schema (Esquema Estrela)
- Exemplo de tabela de fatos
  - Vendas

Id do Tempo	Id do Produto	Id do Funcionário	Unidades Vendidas	Valor de Venda
031011	101	200	10	500
041011	101	200	13	650
051011	101	200	15	700
061011	101	200	20	1000

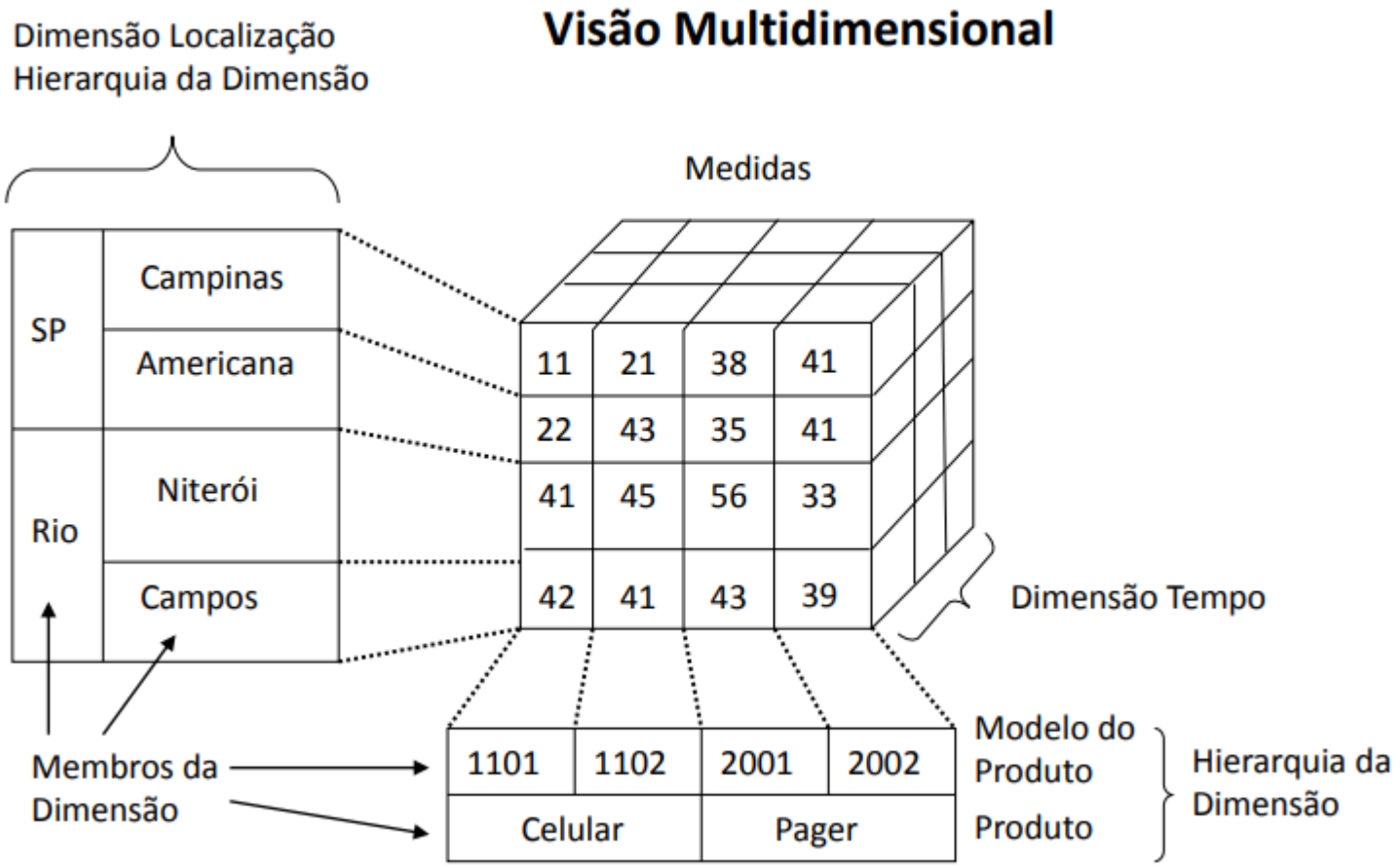
Modelagem Multidimensional

- Membros de uma dimensão
  - São os elementos das dimensões
  - Hierarquia de dimensão
    - Ex.: Cidade, estados e regiões formam a dimensão “Local”, assim como ano, trimestre e mês formam a dimensão “Tempo”



Modelagem Multidimensional

- Membros de uma dimensão



Modelagem Multidimensional

- As dimensões representam entidades que evoluem com o tempo..
  - Por exemplo, um cliente pode deixar de ser solteiro e casar-se
  - Para tratar essas atualizações, pode-se tratar as dimensões de três formas diferentes
  - De acordo com a importância de se ter informações históricas!

Modelagem Multidimensional

- Dimensão Tipo 1
  - O histórico não é relevante!
  - As alterações podem ser feitas diretamente no registro em questão sem salvar o valor anterior
  - Ex.: Godofredo tinha seu estado civil solteiro até 02/07/2013
  - Godofredo casou-se dia 02/07/2013
  - Godofredo teria seu estado civil atualizado para casado

Modelagem Multidimensional

- Dimensão Tipo 1

Id do Cliente	Nome	Estado_Civil
101	Godofredo	Solteiro

Id do Cliente	Nome	Estado_Civil
101	Godofredo	Casado

Modelagem Multidimensional

- Dimensão Tipo 2
  - O histórico é relevante!
  - Inserção de um novo registro na mesma entidade dimensional refletindo a mudança
  - Ex.: Existirão dois registro do Godofredo, o 1º referente a seu estado civil até 02/07/13 e o outro após essa data como casado
  - Na tabela de fatos vendas, o primeiro registro de Godo está vinculado às vendas anteriores a 02/07/13 e o outro às vendas posteriores

Modelagem Multidimensional

- Dimensão Tipo 2

Id do Cliente	Nome	Estado_Civil	Status
101	Godofredo	Solteiro	Antigo
101	Godofredo	Casado	Atual

Modelagem Multidimensional

- Dimensão Tipo 3
  - O histórico é relevante e deseja-se analisar dados usando os status original e atual
  - São necessários campos para armazenar
    - Status original do atributo
    - Status atual do atributo
    - Data efetiva da última alteração do campo
  - Apenas dois status podem ser rastreados: o atual e o original!

Modelagem Multidimensional

- Dimensão Tipo 3

Id do Cliente	Nome	Estado_Civil_Original	Estado_Civil_Atual	Data_Efetiva
101	Godofredo	Solteiro	Casado	02/07/2013

- E se a esposa de Godo largá-lo???

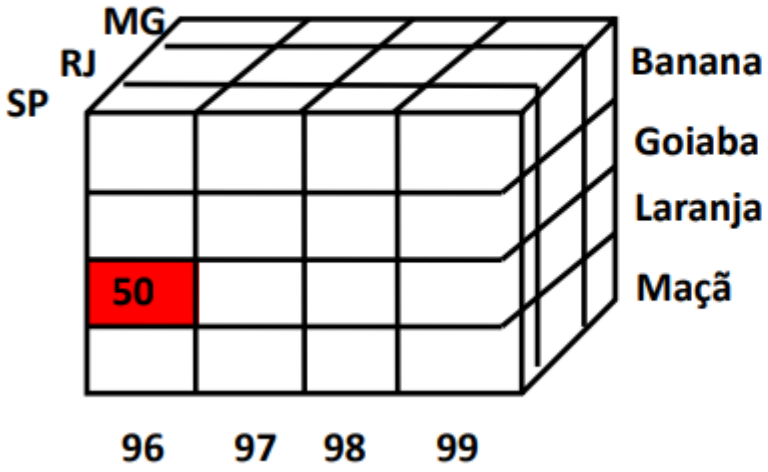
Id do Cliente	Nome	Estado_Civil_Original	Estado_Civil_Atual	Data_Efetiva
101	Godofredo	Solteiro	Divorciado	04/07/2013

- Não terá como fazer análise sobre informações de quando Godo era casado

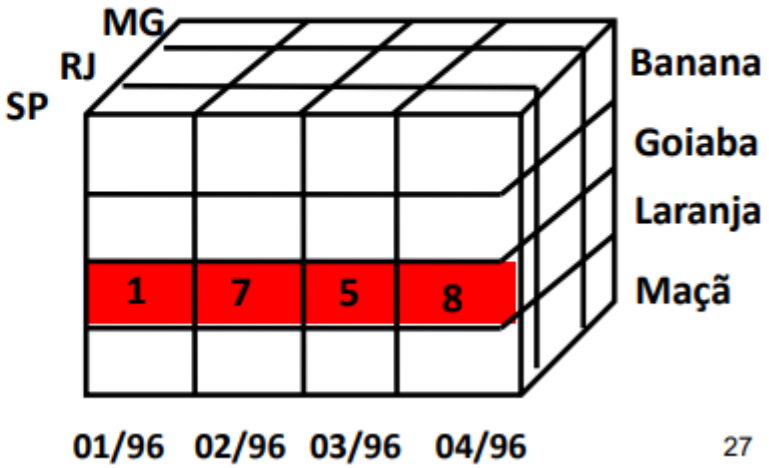
Modelagem Multidimensional

- Granularidade
  - É o nível de detalhe das tabelas
  - Quanto menor o “grão” maior a granularidade

MENOR Granularidade

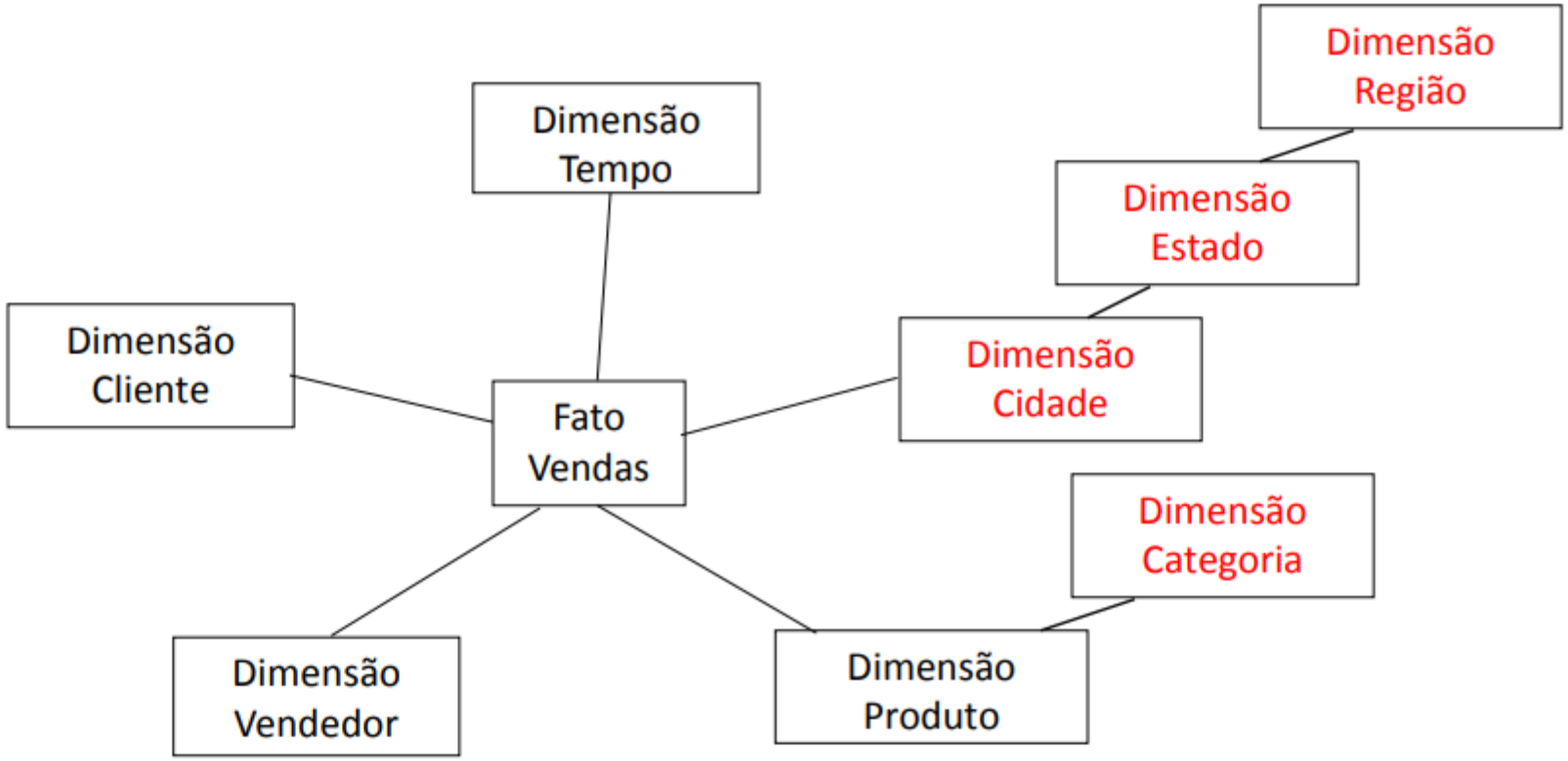


MAIOR Granularidade



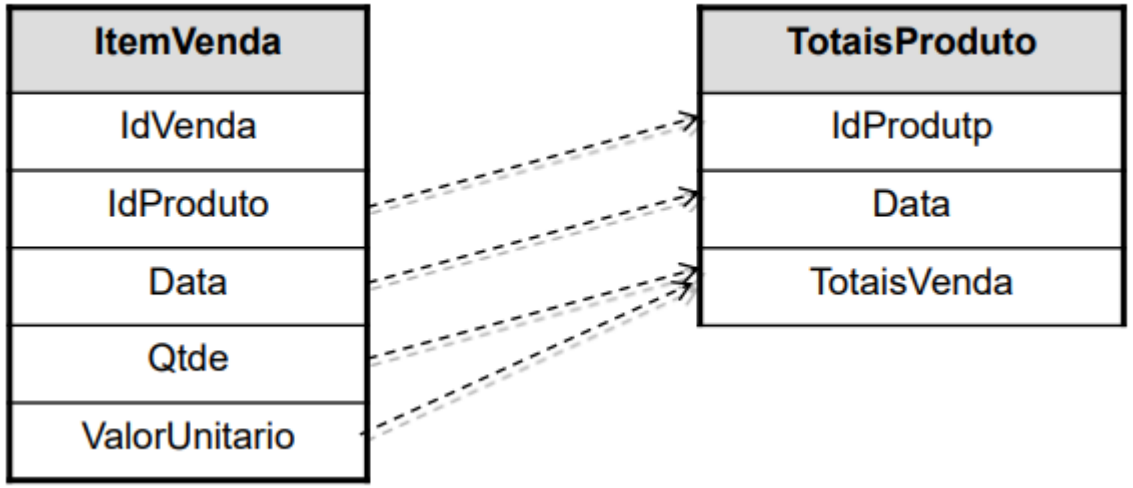
Modelagem Multidimensional

- Esquema Snowflake (Bloco de neve)



Modelagem Multidimensional

- Agregação
  - Através da agregação cria-se novas entidades contendo dados sumarizados



Modelagem Multidimensional

- Quatro passos básicos
  - 1º Definir o FATO de negócio
  - 2º Definir a GRANULARIDADE utilizada
  - 3º Definir as DIMENSÕES do fato
  - 4º Definir as MEDIDAS do fato

Modelagem Multidimensional

- Exemplo
  - Uma rede de restaurantes tem 50 filiais localizadas em vários estados da federação. Cada filial oferece mais de 1000 produtos diferentes nas categorias bebidas e pratos.
  - A diretoria da empresa deseja analisar as vendas, os custos e os lucros obtidos bem como os funcionários mais ativos.
  - Promoções e festivais são utilizados para atrair clientes e potencializar as vendas.

Modelagem Multidimensional

- Exemplo
  - A diretoria da empresa determinou que é estratégico para a tomada de decisões analisar o **movimento diário de cada produto**, para que possa direcionar as promoções ou festivais de acordo com os resultados das análises realizadas
  - Avaliar o movimento diário de cada produto consiste em analisar as **vendas** de produtos, levando em conta os **preços praticados** e as **filiais** que realizaram tais vendas

Modelagem Multidimensional

- Exemplo
  - 1º Definir o FATO
  - Qual elemento central a empresa deseja analisar???
  - R.: Vendas

Modelagem Multidimensional

- Exemplo
  - 2º Definir a GRANULARIDADE
  - Em que nível de detalhe a empresa deseja analisar???
    - “é estratégico para a tomada de decisões analisar o movimento diário de cada produto”
  - R.: Diário
    - (com respeito a outras dimensões não foi especificado)

Modelagem Multidimensional

- Exemplo
  - 3º Definir as DIMENSÕES
    - Quais aspectos são relevantes para se realizar as análises que a empresa solicita do fato???
    - De forma geral, alguns fatores a se observar são
      - O quê ⟶ Produto
      - Quem ⟶ Funcionário
      - Quando ⟶ Tempo
      - Onde ⟶ Local (filial)
    - Pode-se ainda levar em conta outros objetivos especificados para a análise dos fatos, como Promoções

Modelagem Multidimensional

- Exemplo
  - 4º Definir as MEDIDAS
    - Como o desempenho de vendas pode ser medido???
    - R.: Quantidade vendida, valor unitário e total da venda, valor da compra

Modelagem Multidimensional

- Exemplo – Modelo estrela correspondente