

DS340 – Banco de Dados 3

Data Warehouse e Modelo Dimensional

Prof. João Eugenio Marynowski
jeugenio@ufpr.br



■ SGBDs Paralelos

- Arquiteturas: Memória, Disco e Rede (Nothing)
- Tipos de paralelismo: Intra-op, Inter-op e Inter-query
- Particionamento: Range, Hash e Round Robin

■ SGBDs **Distribuídos**

- Tipos: Homogêneos e Heterogêneos/Múltiplos
- Arquiteturas: Centralizada e **Distribuída**
- Replicação: total/parc, sinc/assinc
- Fragmentação: horizontal, vertical e mista
- Alocação: tamanho, transações, operações desconectadas
 - **Disponibilidade, Confiabilidade e Desempenho**
- Níveis de Transparência
 - falhas, desempenho, heterogeneidade, rede/distribuição
nomes, fragmentação, replicação e localização



Atividade SGBDD

- Dado o estudo de caso de uma empresa de entregas X com sede em Curitiba e uma filial em SP, escolha um SGBD de sua preferência e descreva sobre o suporte deste SGBD para implementar um BDD para esta empresa. O BDD deve ser formado pelas tabelas Empregado e Entrega, que contém informações dos empregados e entregas, respectivamente, de forma que as tabelas sejam **fragmentadas** e **replicadas** para aumentar a disponibilidade, confiabilidade e desempenho. A tabela Empregado é replicada nas duas cidades. A tabela Entrega é fragmentada de acordo com o endereço e é replicada de forma assíncrona na outra cidade. Descrever os comandos e configurações necessárias para implementar este estudo de caso.



Não!!!

SQL File 3* SQL File 5* Administration - Data Export x

Root
Data Export

Object Selection Export Progress

Tables to Export

Exp...	Schema
<input type="checkbox"/>	bd3
<input checked="" type="checkbox"/>	replica
<input type="checkbox"/>	sys

Exp... Schema Objects

Refresh Dump Structure and Dat v Sele

Objects to Export

☐ Dump Stored Procedures and Functions ☐ Dump Events

Export Options

☒ Export to Dump Project Folder C:\Users\gusta\Documents\dumps\Dump20190813

Each table will be exported into a separate file. This allows a selective restore, but may be slower.

☐ Export to Self-Contained File C:\Users\gusta\Documents\dumps\Dump20190813.sql

All selected database objects will be exported into a single, self-contained file.

SQL File 3* SQL File 5* Administration - Data Import/Res... x

Root
Data Import

Import from Disk Import Progress

Import Options

☒ Import from Dump Project Folder C:\Users\gusta\Documents\dumps ...

Select the Dump Project Folder to import. You can do a selective restore.

Load Folder Contents

☐ Import from Self-Contained File C:\Users\gusta\Documents\dumps\export.sql ...

Select the SQL/dump file to import. Please note that the whole file will be imported.

Default Schema to be Imported To

Default Target Schema: v New...

The default schema to import the dump into.
NOTE: this is only used if the dump file doesn't contain its schema, otherwise it is ignored.

Select Database Objects to Import (only available for Project Folders)

Imp...	Schema
--------	--------

Imp...	Schema Objects
--------	----------------



Oracle/PostgreSQL/SQL Server... Mysql...

- DATAFILE
- DATABASE LINK
- Empregado@Curitiba
- TRIGGER ... *@Curitiba ... *@SaoPaulo
 - https://wiki.postgresql.org/wiki/Replication,_Clustering,_and_Connection_Pooling
- Replication Manager
 - RELAY LOG
 - GoldenGate (<https://www.oracle.com/technical-resources/articles/database/datawarehouse-sqlserver-goldengate.html>)
- Cluster
- Streams

Oracle Homogêneo

- A integração é basicamente toda baseada em Oracle Database Links
- Database Administrator's Guide / 31 Distributed Database Concepts
 - <https://docs.oracle.com/en/database/oracle/oracle-database/12.2/admin/distributed-database-concepts.html#GUID-36A35EC6-D3BA-4284-ADA0-732E10F3425E>
- Database Administrator's Guide / 31.2 Database Links
 - <https://docs.oracle.com/en/database/oracle/oracle-database/12.2/admin/distributed-database-concepts.html>
- Database Administrator's Guide / 31.2.9 Creation of Database Links: Examples
 - <https://docs.oracle.com/en/database/oracle/oracle-database/12.2/admin/distributed-database-concepts.html#GUID-A1B3AEB2-5625-4ABF-9378-E941DC0DB2A0>



Oracle Heterogêneo/Multiplus

- Oracle Database Gateway e ODBC
- Database Administrator's Guide / 31 Distributed Database Concepts
 - [<https://docs.oracle.com/en/database/oracle/oracle-database/12.2/admin/distributed-database-concepts.html#GUID-36A35EC6-D3BA-4284-ADA0-732E10F3425E>]
- Heterogeneous Connectivity User's Guide / 3 Features of Oracle Database Gateways
 - <https://docs.oracle.com/en/database/oracle/oracle-database/12.2/heter/features-of-oracle-database-gateways.html#GUID-13FD7B76-9463-4EE3-8DDA-13C1AA7C815C>
- Installation and Configuration Guide for IBM AIX on POWER Systems (64-Bit), Linux x86-64, Oracle Solaris on SPARC (64-Bit), Oracle Solaris on x86-64 (64-Bit), and HP-UX Itanium
 - <https://docs.oracle.com/en/database/oracle/oracle-database/19/otgis/toc.htm>
- User's Guide / Oracle Database Gateway for ODBC
 - <https://docs.oracle.com/en/database/oracle/oracle-database/19/odbcu/toc.htm>

Assuntos Passados (2)

- Sistemas de Informações Geográficas (SIG)
- Dados espaciais e geográficos
- BD Geográficos (BDG) e SGBDs Geográficos (SGBDG)
 - Forma, Localização na Terra e a sua validade
 - Modelo de Campo, de Objeto e de Rede (Grafo)
 - Atualização, Análise, Consulta e Visualização de DG
 - Operações de Transformações, Topológicas, Métricas e Estatísticas
 - SQL estendido
 - Visualização em mapa georreferenciado



Atividade SGBDG

- Considerando um SGBD de sua preferência, descreva sobre o suporte deste SGBD para diferentes modelos de dados geográficos, operações e formas de visualização que sejam necessários para o armazenamento e a identificação da localização dos acessos ao SEU sistema web. Indique as extensões e comandos necessários para se obter as regiões geográficas que mais tiveram acesso ao sistema.

- IP -> Coordenada geográfica
 - IP-API, GeoLite, GeoNames

- Atributos alfanuméricos

- POINT, LINE/LINESTRING, POLYGON (GEOGRAPHY), ...

- DISTANCE, AREA, CONTAINS, TOUCHES, INTERSECTS, ...



Análise e Desenvolvimento
de Sistemas



PARTE 2

Data Warehouse e Modelo Dimensional



Dado e a Tomada de Decisão

Sabedoria | Decisão



COMPRA!

Tendência de alta

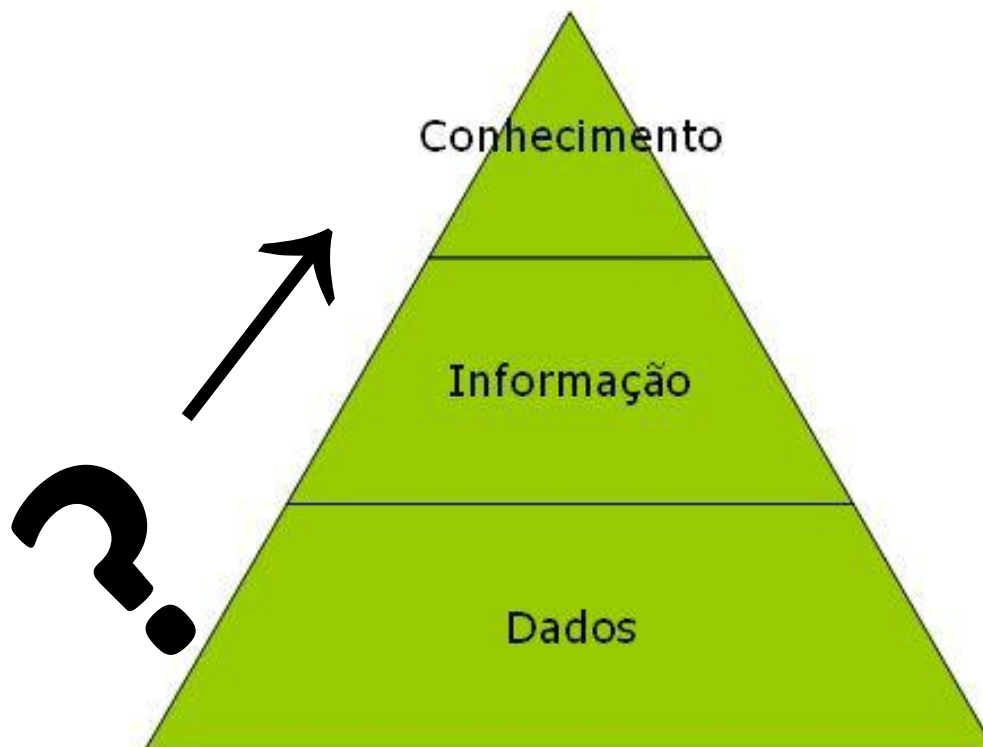
35000 pontos

35000



Dado e a Tomada de Decisão

Sabedoria | Decisão



COMPRA!

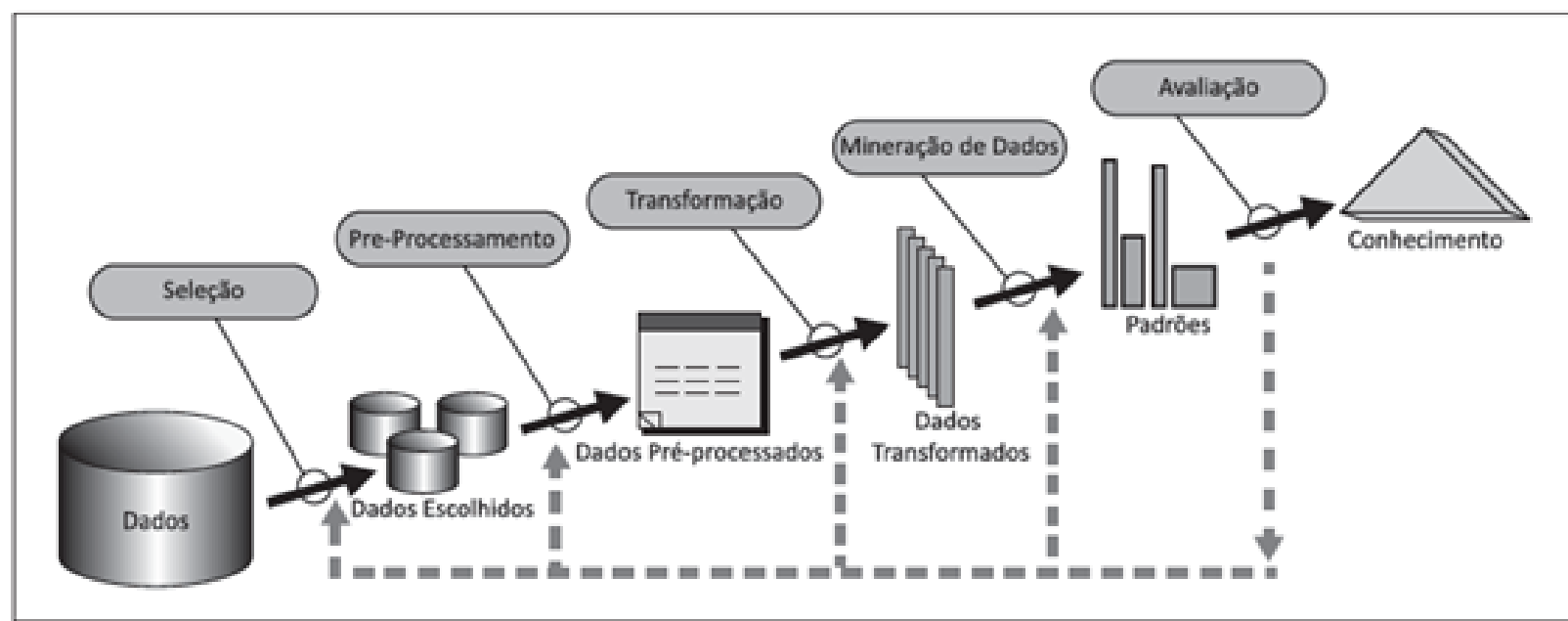
Tendência de alta

35000 pontos

35000



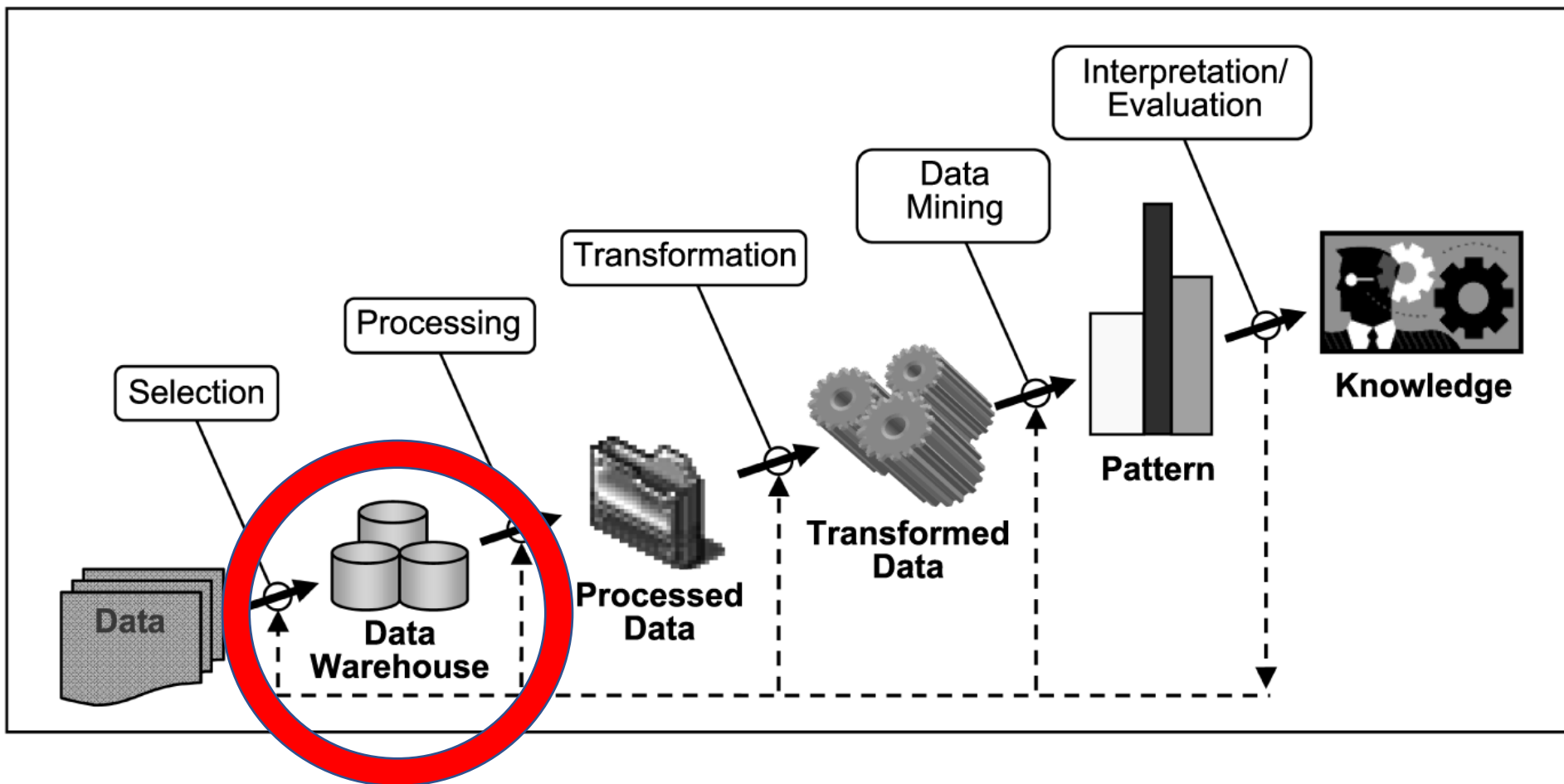
Processo de Descoberta de Conhecimento KDD - Knowledge Discovery in Databases



- “É um processo, de várias etapas, não trivial, interativo e iterativo, para identificar padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados” (Fayyad et al., 1996)



Processo de Descoberta de Conhecimento KDD - Knowledge Discovery in Databases



(Fayyad et al., 1996)

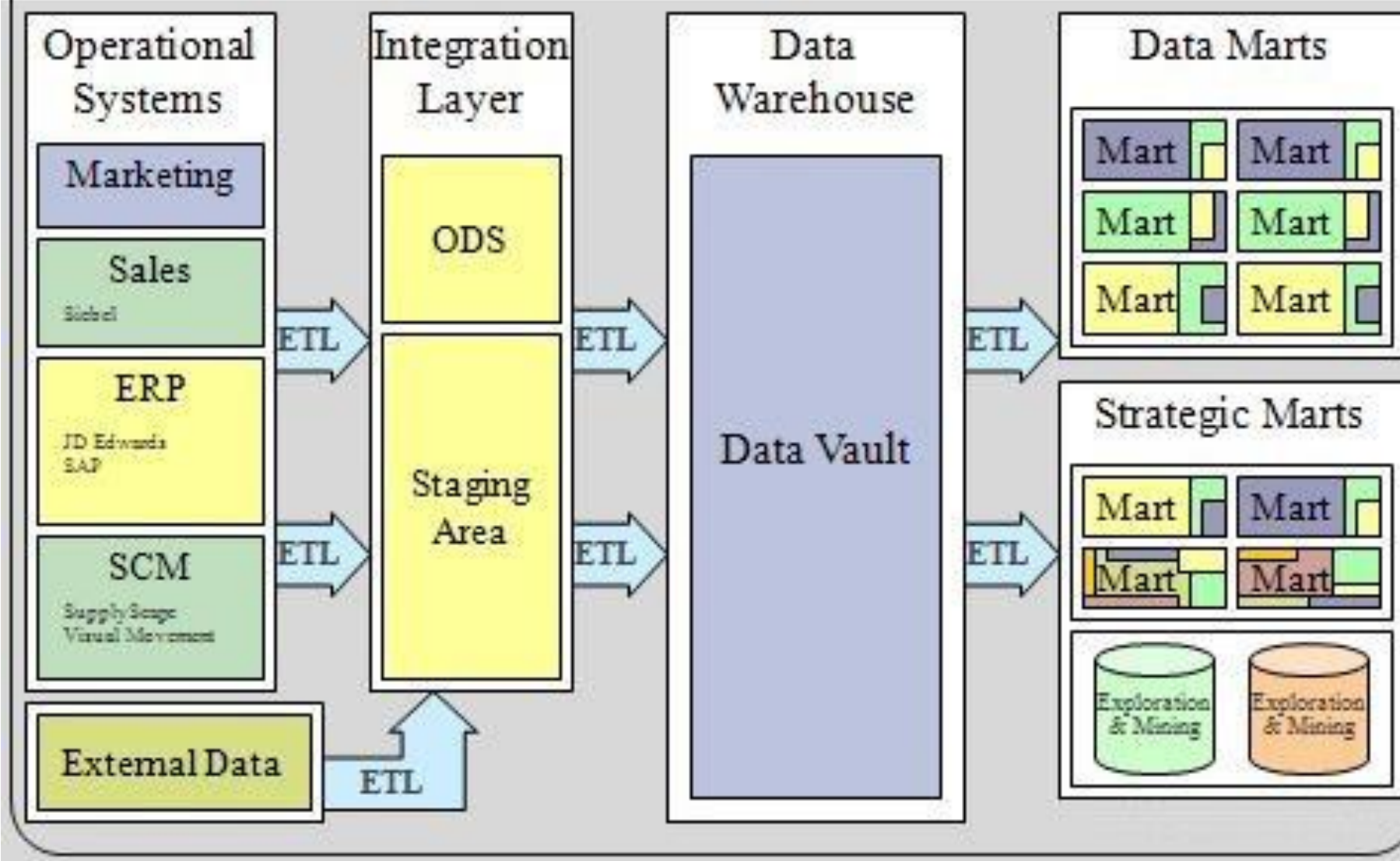


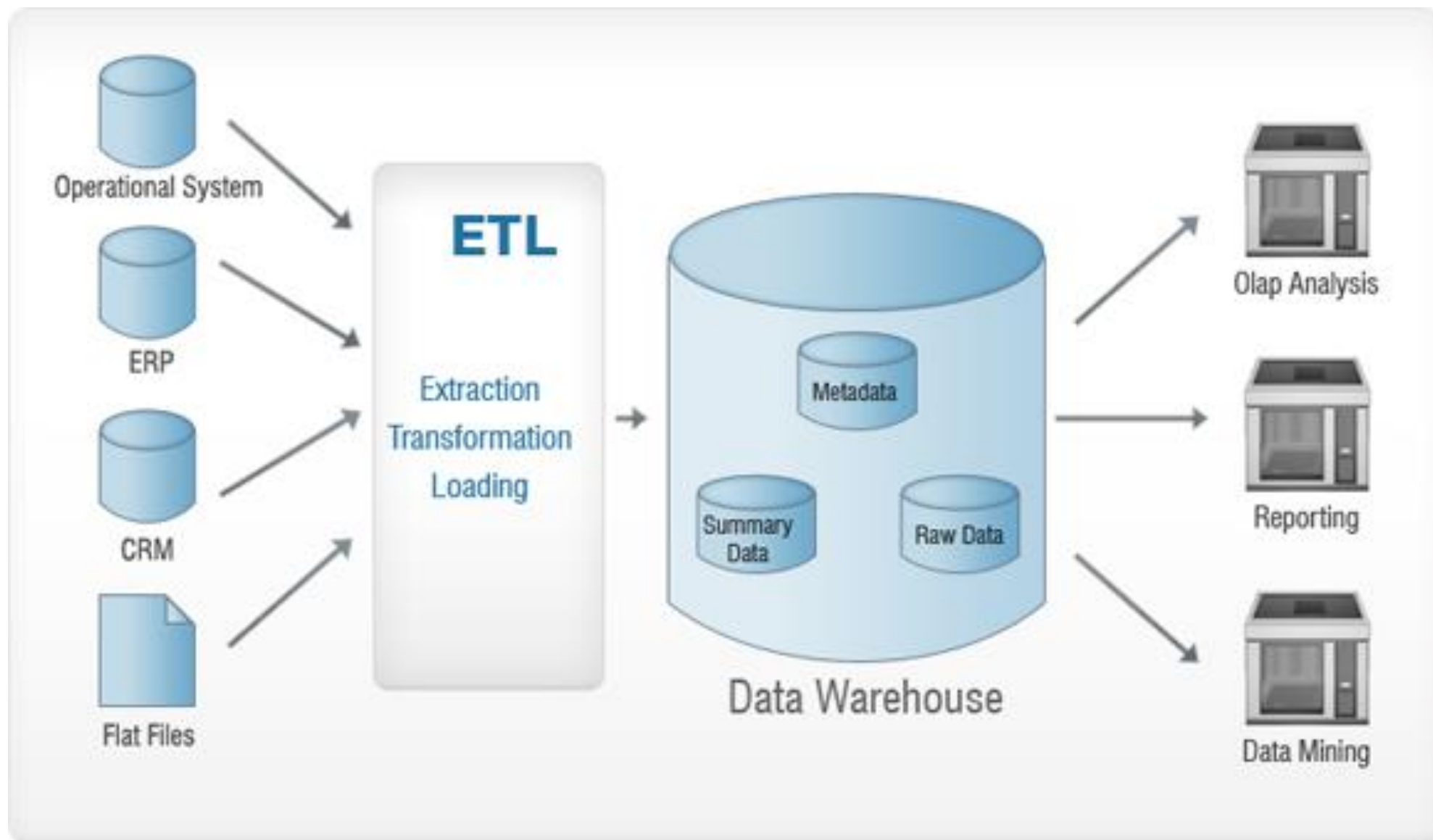
Data Warehouse (DW)

- Armazém ou depósito de dados
- [Inmon'1992]
 - “Uma coleção de dados orientada a assunto, integrada, variável no tempo e não volátil para o suporte às decisões da gerência.”
- Dados históricos, análise temporal e de tendências
- VLDB (Very Large Data Bases) – TB, PB, ...
- Suporte as soluções de BI
 - KDD, DM, DSS/EIS, OLAP,...
- Modelagem dimensional ou multidimensional
[colunar, chave-valor, grafo, geográfico, documento, ...]



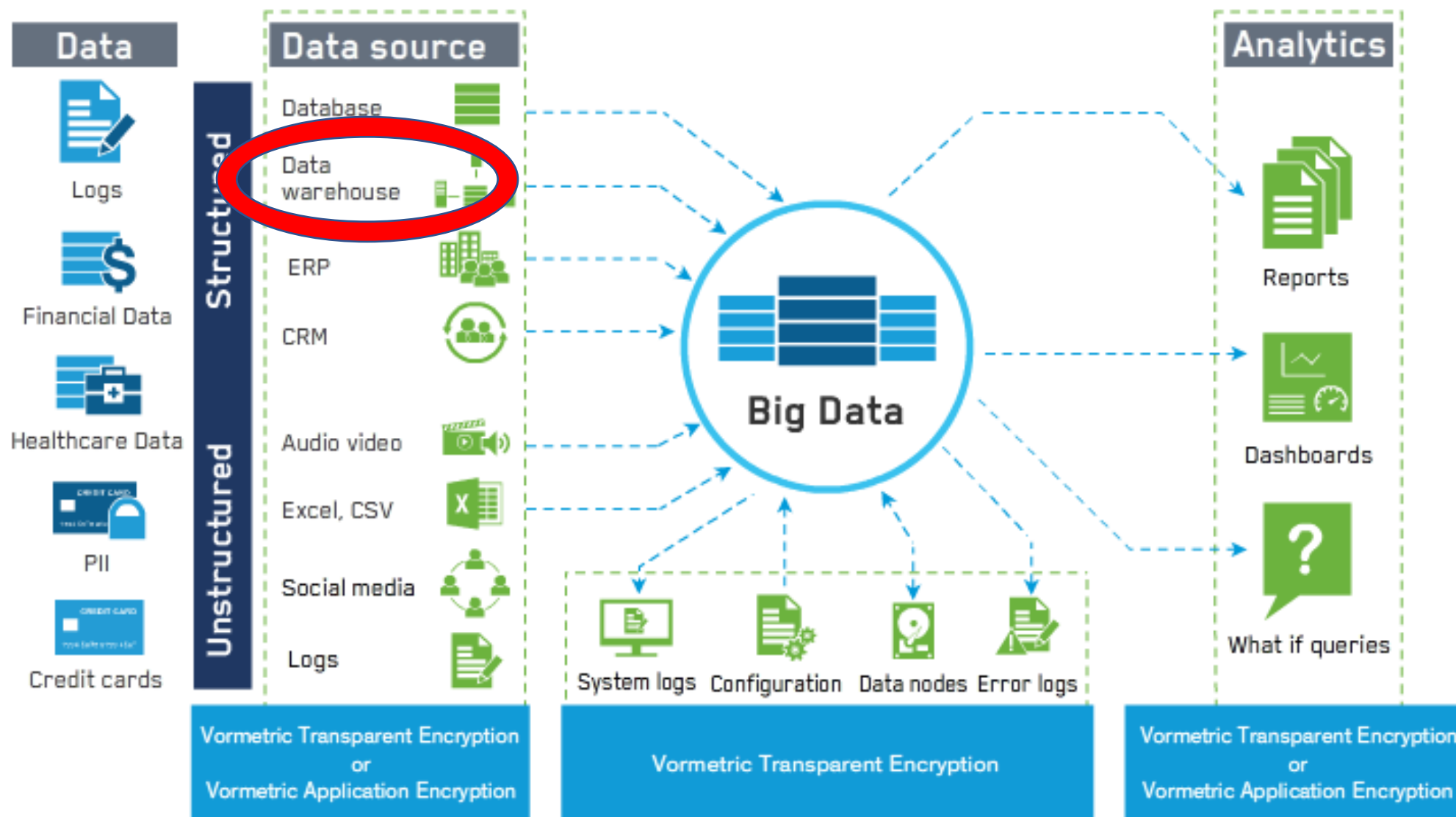
Data Warehouse







Data Lake





BDR x DW

- BDR (Banco de Dados Relacional)
 - Empregados e horas trabalhadas
 - Peças em estoque e quantidade de pedidos
- DW
 - Quantidade de peças com estoque zerado nos últimos 3 anos?
 - Quantidade de entregas efetuadas erroneamente nos últimos 2 anos?
 - Distribuição da renda dos devedores no último ano?

Características de um DW

- Orientado a assuntos
- Integrado
- Variável no tempo
- Não-volátil

Características de um DW

- Orientado a assuntos
 - Núcleo da empresa, análise p/ decisão, não operação diárias(transacionais), visão simples
- Integrado
 - Fontes heterogêneas, limpeza/integração/consistência e conversão
- Variável no tempo
 - Visão histórica, não operacional, tempo presente nas estruturas
- Não-volátil
 - Armazenamento separado da operação



Modelo Dimensional/Multidimensional

- Dados apresentados em duas ou mais dimensões
- Dimensões possibilitam observar os dados sob várias óticas ou perspectivas
- Principais componentes
 - Fatos
 - Medidas
 - Dimensões

Fato e Medidas

- **Fato** é um conjunto de itens de dados quantitativos
- **Medidas** são atributos (numéricos) associados ao fato
 - ex.: quantidade, valor, percentual, média, ...
- Reflete a evolução dos negócios no dia a dia de uma organização
- Escopo ou assunto a ser tratado
- O Fato muda suas medidas com o tempo (é evolutivo)
 - Compra/Consumo
 - Venda
 - Lucro

Dimensão

- Reflete a perspectiva dos fatos
- Possíveis formas de visualizar os dados
- Conjunto de itens de dados qualitativos
- Valor descritivo
 - Tempo (Data, mês, ano)
 - Local
 - Cliente
 - Fornecedor

Fato Vendas

- Vendas totais, classificadas por região

Região	Venda
Sul	\$ 1.500,00
Sudeste	\$ 5.000,00
Nordeste	\$ 2.350,00
Norte	\$ 1.890,00
Centro-Oeste	\$ 1.732,00
Total de Vendas	\$ 11.472,00

- Medida? Valor da venda
- Dimensão? Região (localidade)



Vendas Agrupadas por Região e Trimestre

Região	Trimestre	Venda
Sul		
	1	\$ 250,00
	2	\$ 700,00
	3	\$ 250,00
	4	\$ 300,00
Sudeste		
	1	\$ 750,00
	2	\$ 800,00
	3	\$ 500,00
	4	\$ 300,00

- Medida?.... Dimensões?...
- Como adicionar Grupos de Produtos ?



Vendas por Região, Trimestre e Grupo

Região	Trimestre	Grupo de Produção	Venda
Sul			
	1		
		Enlatados	\$ 50,00
		Cereais	\$ 200,00
	2		
		Enlatados	\$ 300,00
		Cereais	\$ 500,00
	3		
		Enlatados	\$ 25,00
		Cereais	\$ 225,00
	4		
		Enlatados	\$ 200,00
		Cereais	\$ 100,00
Suldeste			



Análise e Desenvolvimento
de Sistemas



PARTE 3

Modelagem Dimensional



Tabela Fato: Ocorrências

Ocorrência	2002	2003	2004
Assalto à mão armada	123	109	158
Furtos em residências	75	90	101
Furtos em veículos	243	250	332
Assassinatos	89	77	167
Estupros	2	24	69

Medida? **Dimensão?**



Abordagem Dimensional

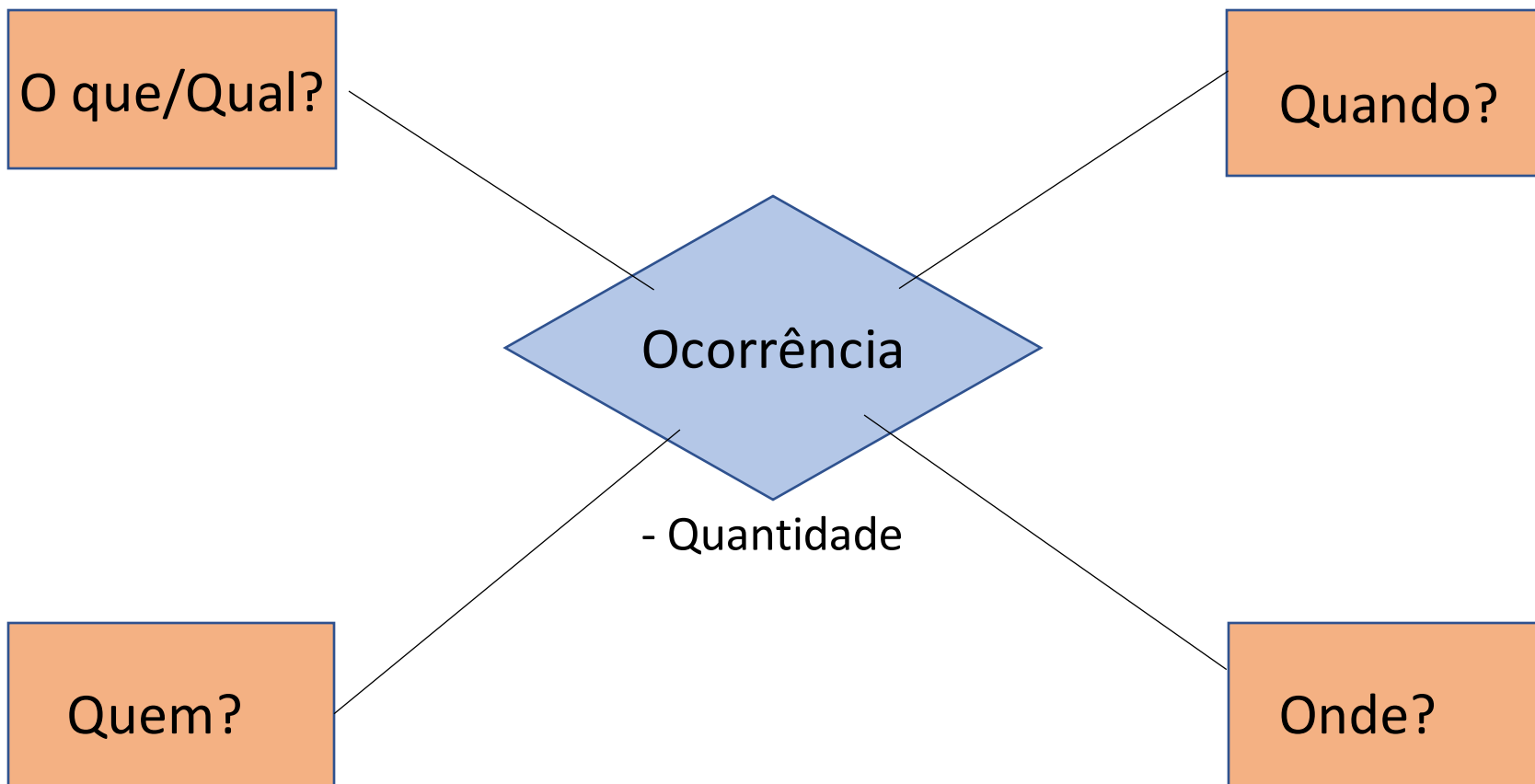


Tabela Fato: Ocorrências

Ocorrência	2002	2003	2004
Assalto à mão armada	123	109	158
Furtos em residências	75	90	101
Furtos em veículos	243	250	332
Assassinatos	89	77	167
Estupros	2	24	69

Medida: Quantidade

Dimensão: Ocorrência(Tipo) e Ano

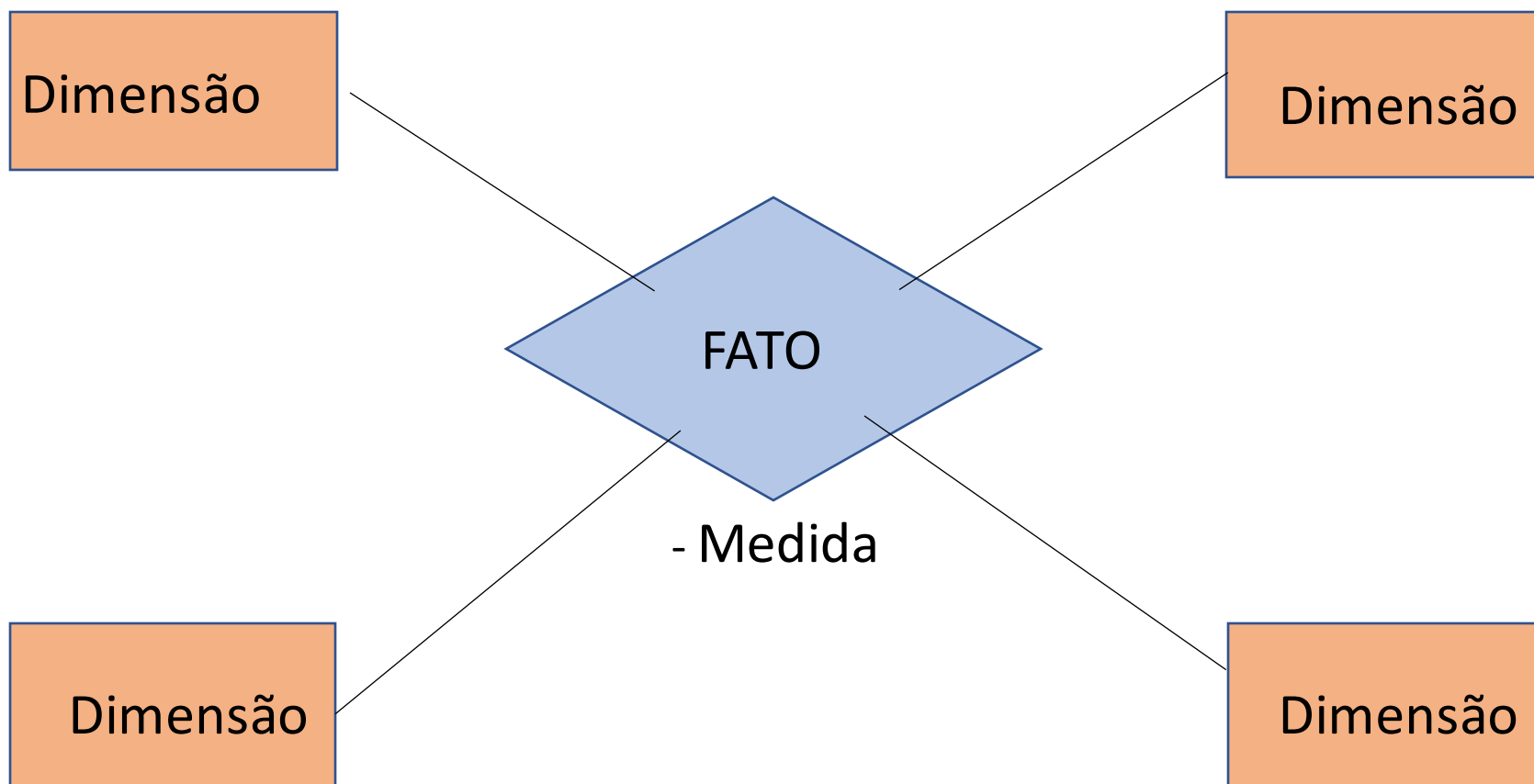


Fato e três Dimensões

Ocorrência	2002	2003	2004	PR	SC	RS
Assalto à mão armada	123	109	158	47	31	77
Furtos em residências	75	90	101	27	...	34
Furtos em veículos	243	250	332	97
Assassinatos	89	77	167	49
Estupros	2	24	69	17



Modelo Dimensional





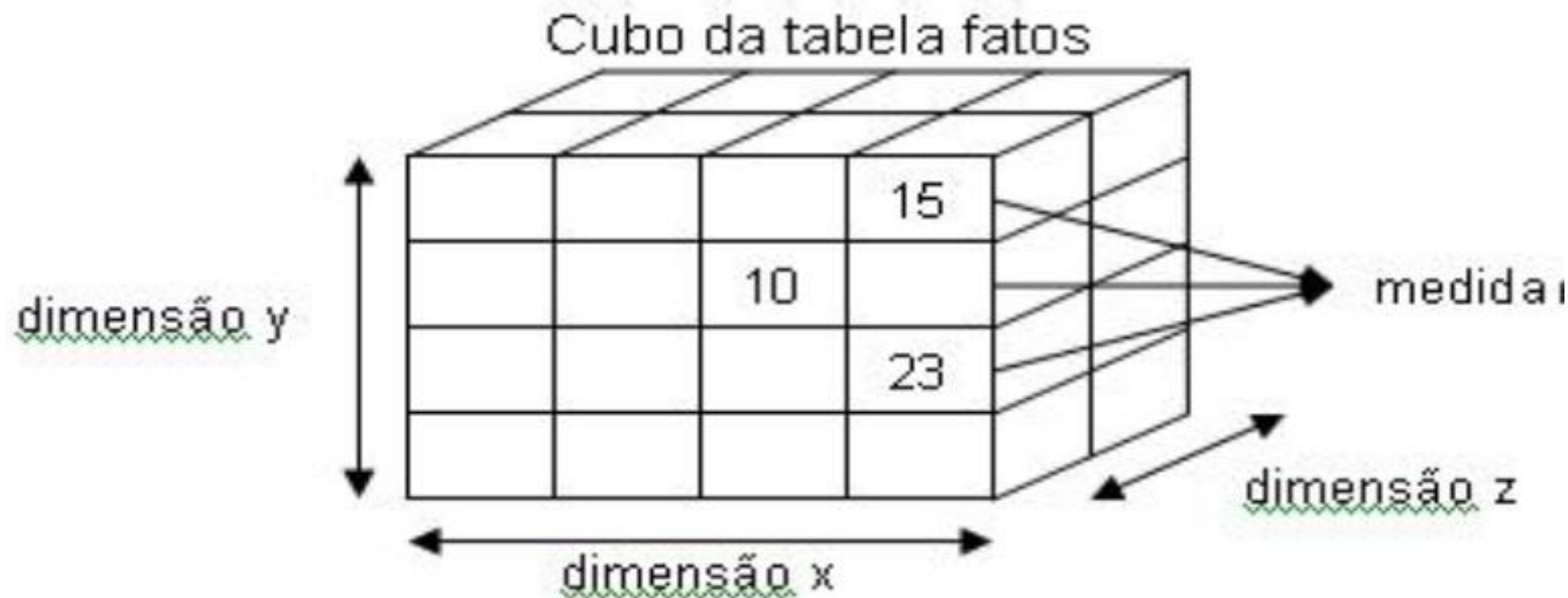
Fato e três Dimensões

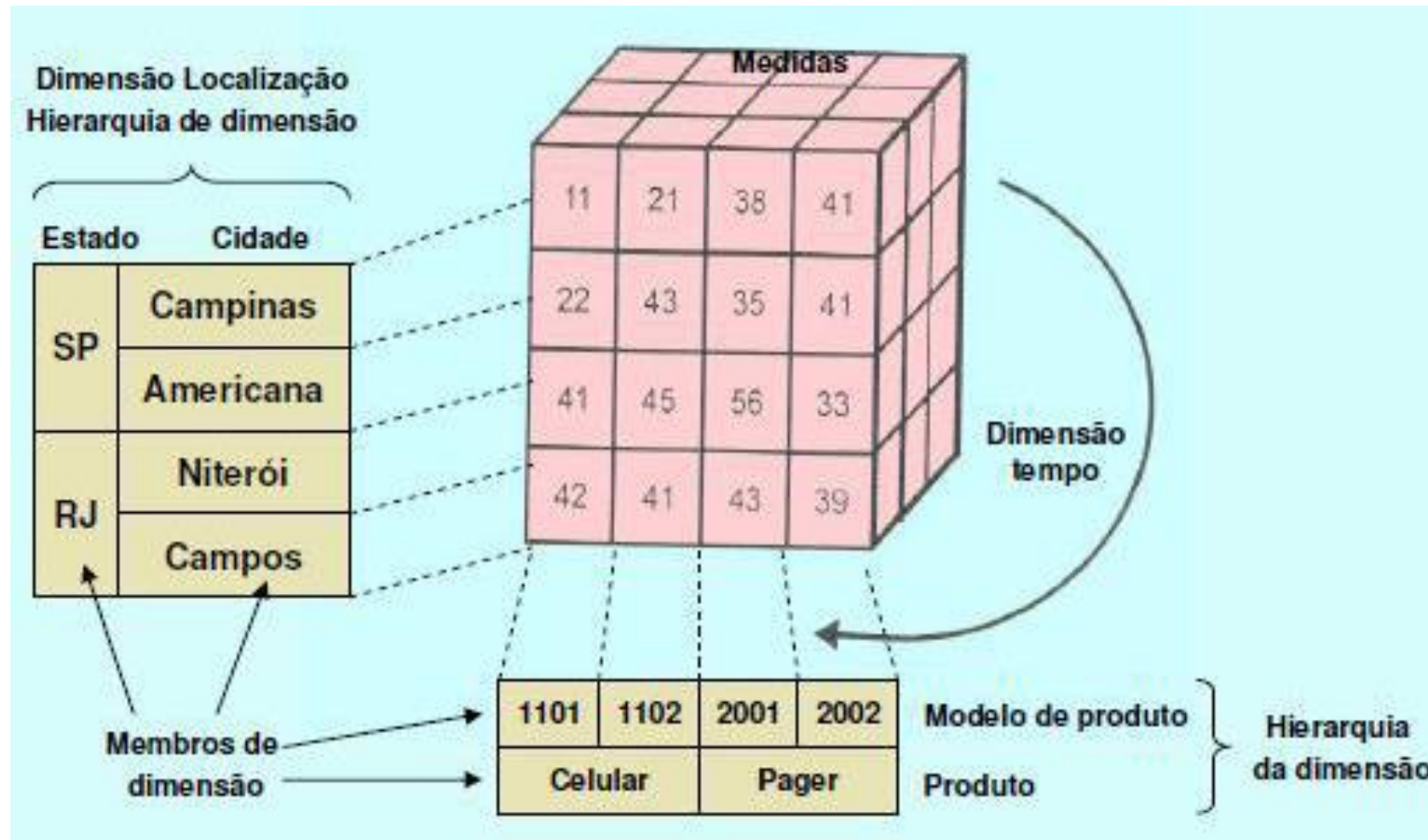
Ocorrência	2002	2003	2004	PR	SC	RS
Assalto à mão armada	123	109	158	47	31	77
Furtos em residências	75	90	101	27	...	34
Furtos em veículos	243	250	332	97
Assassinatos	89	77	167	49
Estupros	2	24	69	17



Cubo ou Hipercubo

Representação do Modelo Dimensional

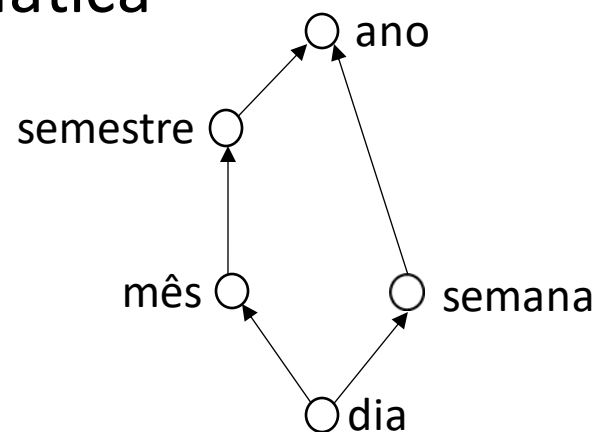




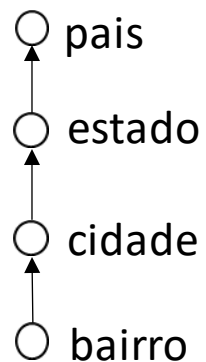


Hierarquia de Dimensões

- Uma dimensão pode conter muitos membros
- Hierarquia esquemática

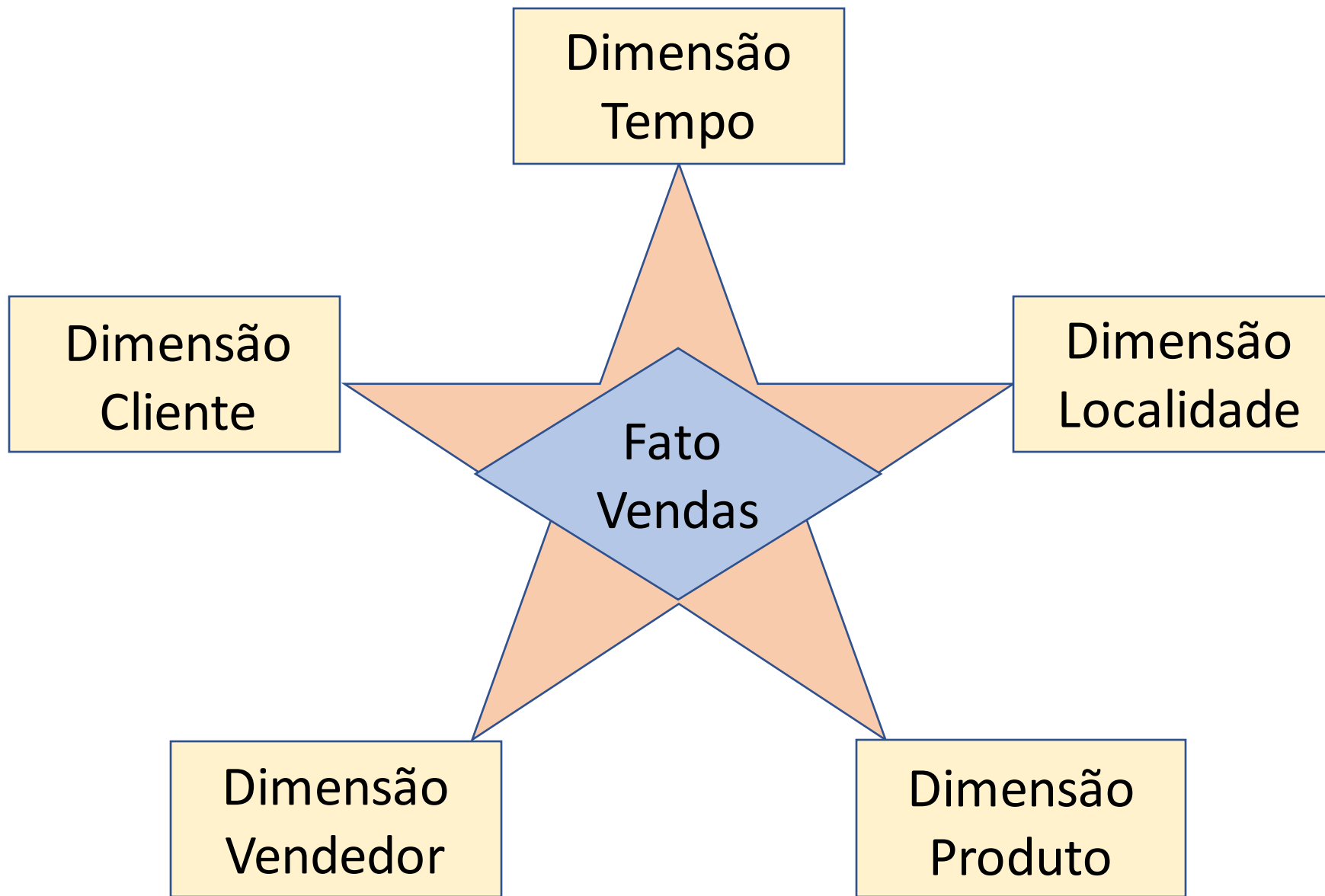


- Hierarquia de agrupamento



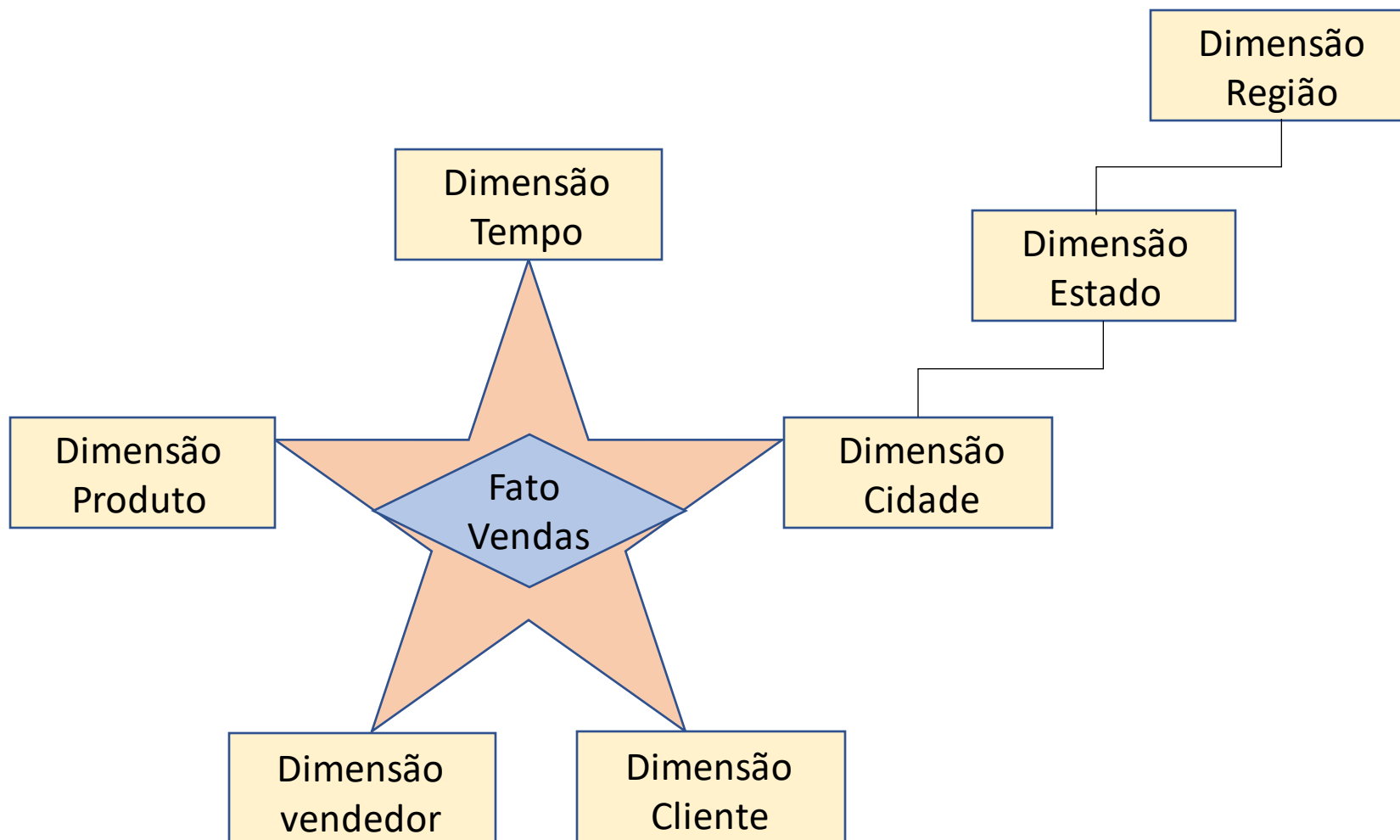


Esquema Estrela Star Schema



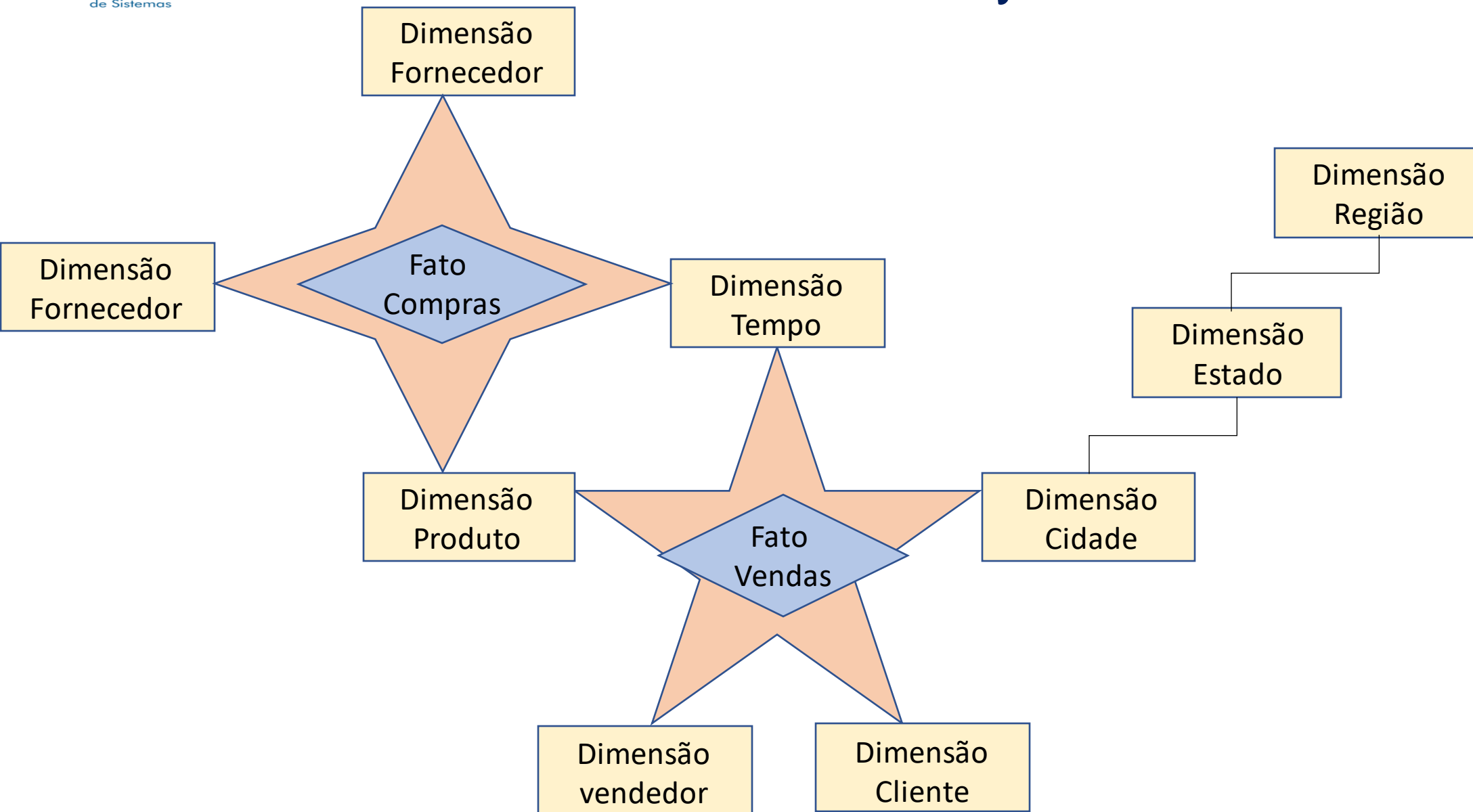


Esquema Flocos de Neve Snow Flake Schema



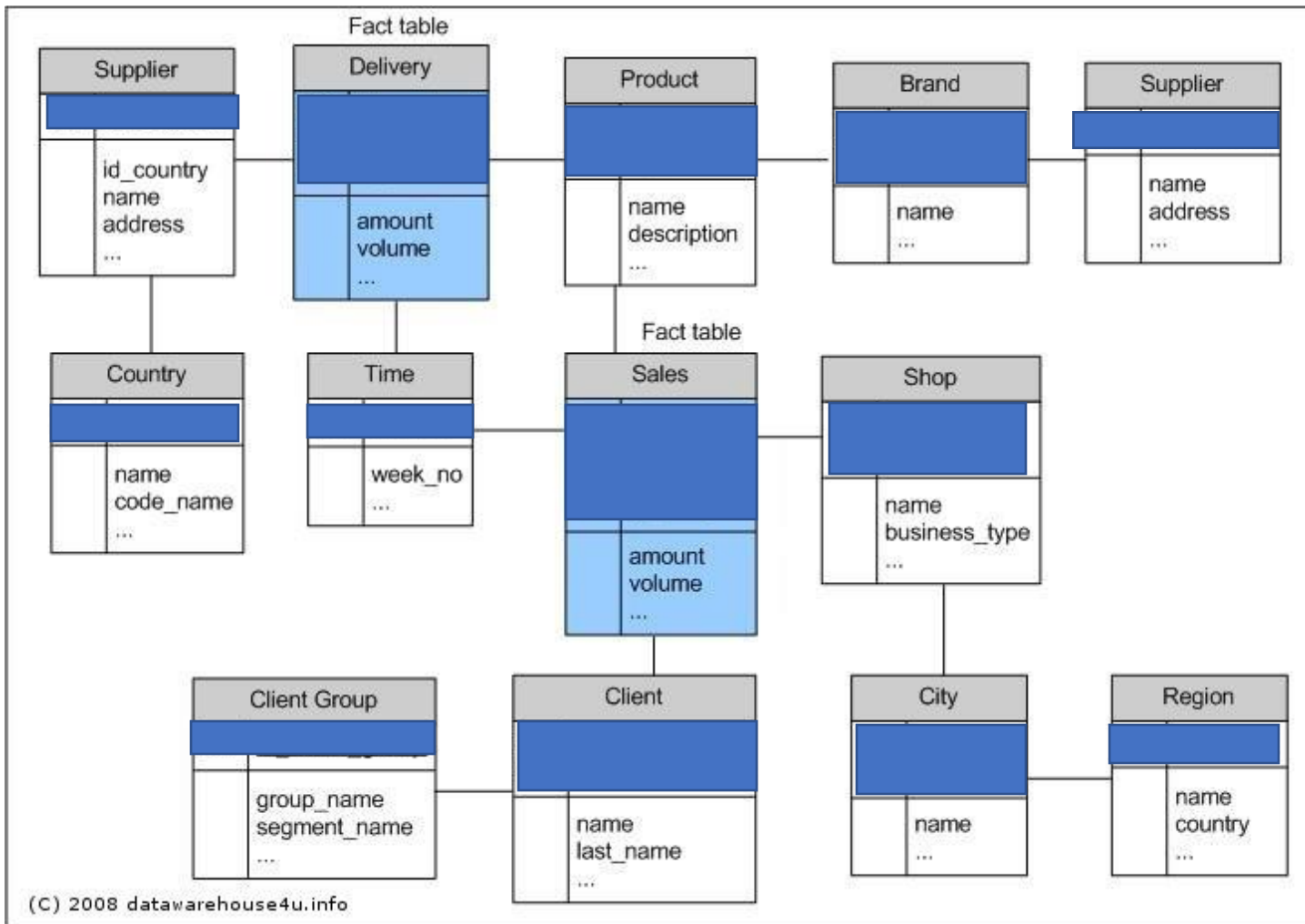


Esquema Constelação/Galáxia Constellation/Galaxy Schema





Esquema Constelação/Galáxia Constellation/Galaxy Schema



Questões

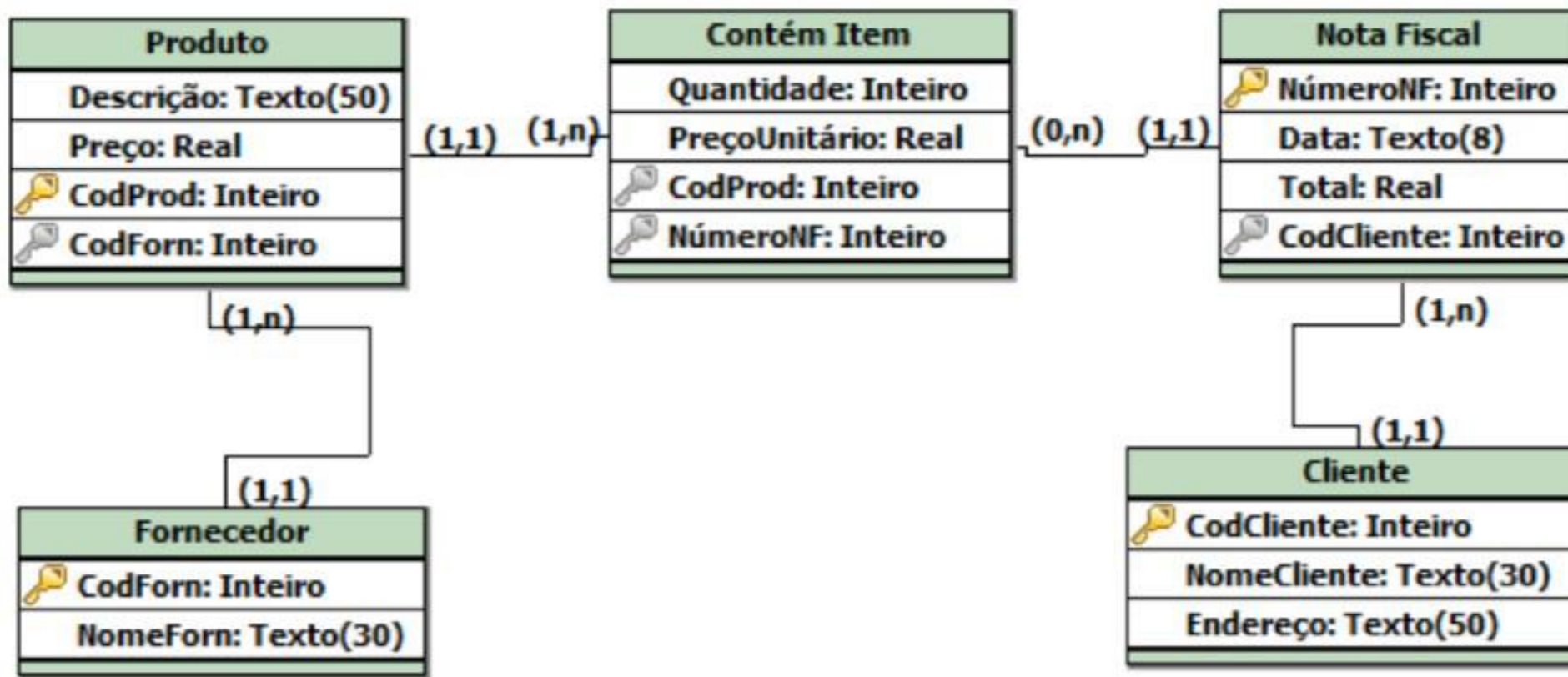
- Qual melhor?
 - Redução de espaço
 - Inconsistências
 - Junções
 - ...

Modelagem Dimensional

- 1) Definir o fato/área/assunto de negócio
- 2) Definir as medidas (dados quantitativos/indicadores)
- 3) Definir as dimensões (dados qualitativos)
 - Quatro perguntas básicas (?)
- 4) Definir o esquema empregado
 - Hierarquia de dimensões ou hierarquias múltiplas
 - Granularidade/agrupamento desejado para os dados
 - Considerar os volumes e dificuldades de se obter o nível desejado



Diagrama Relacional – Nota Fiscal



Qual modelo dimensional? Ou quais?

Modelagem Dimensional

- 1) Definir o fato/área/assunto de negócio
- 2) Definir as medidas (dados quantitativos/indicadores)
- 3) Definir as dimensões (dados qualitativos)
 - Quatro perguntas básicas (?)
- 4) Definir o esquema empregado
 - Hierarquia de dimensões ou hierarquias múltiplas
 - Granularidade/agrupamento desejado para os dados
 - Considerar os volumes e dificuldades de se obter o nível desejado



Atividade

- Dado o Diagrama Relacional – Nota Fiscal, apresente um diagrama conceitual (fato(s), medida(s) e dimensões) para um BD Dimensional com o assunto vendas e que responda as seguintes consultas gerenciais
 - Quantidade de vendas por produto ao longo dos meses, independente dos anos
 - Valor vendido por cliente ao longo dos meses de cada ano
 - Quantidade de vendas por cliente ao longo dos meses, independente dos anos



Análise e Desenvolvimento
de Sistemas

Próximo Assunto



- Modelagem dimensional e Projeto de BDM