

// MARÇO 2021

DATA WAREHOUSING

Aplicações sobre Base de Dados

Elaborado por:

Filipa Mata, Guilherme Inocêncio e Rui Carrilho

TABLE OF CONTENTS



01

CONCEITOS

DW, Data Mart, ETL

02

MODELO DE DADOS

03

OLAP

OLAP vs OLTP, MOLAP,
ROLAP, HOLAP e esquemas

04

MINERAÇÃO DE DADOS

Definição e Técnicas



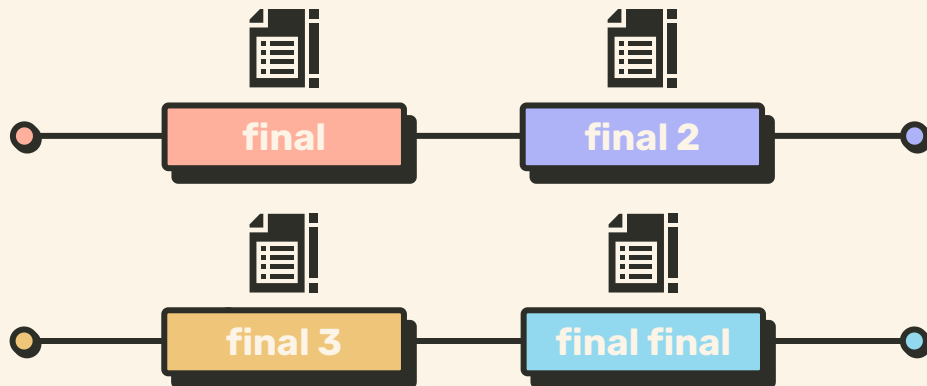
01

CONCEITOS

DATA WAREHOUSE

Armazenamento e Análise de um elevado número de dados

- Centralização de dados provenientes de diversas fontes.
- Facilitam a consulta dos dados.



PRINCIPAIS CARACTERÍSTICAS



SUBJECT-ORIENTED

Informação organizada por temáticas



VARÍÁVEL COM O TEMPO

Os dados são relativos a um determinado tempo

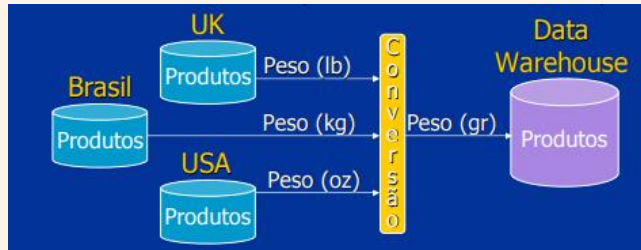


PRINCIPAIS CARACTERÍSTICAS



INTEGRATED

Formato dos dados é padronizado



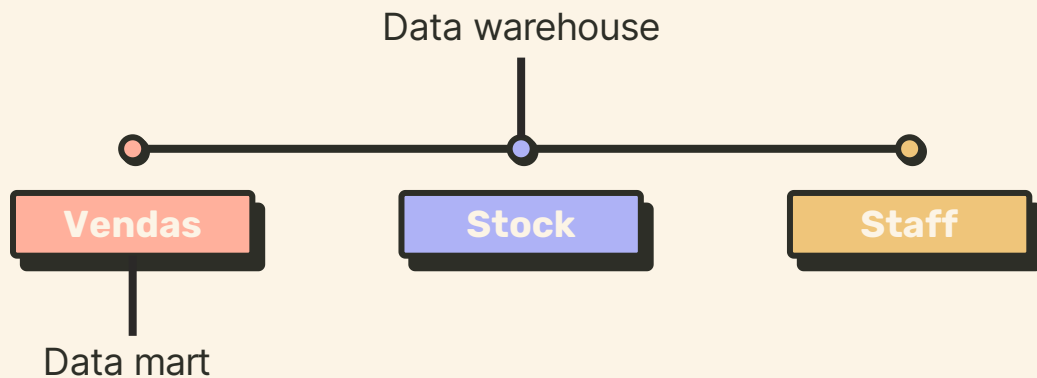
NÃO-VOLÁTIL

Os dados armazenados não se alteram



DATA MART

Subconjunto de dados
provenientes de um Data
Warehouse



ETL



E

T

L



EXTRACT

Extração de dados

TRANSFORMATION

Transformação de
acordo com o negócio

LOAD

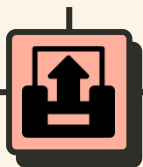
Carregamento dos
dados

ETL



EXTRAÇÃO

De dados de diversos sistemas



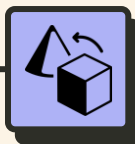
CARREGAMENTO

Para um Data Mart ou Data Warehouse



TRANSFORMAÇÃO

Conforme as regras de negócio

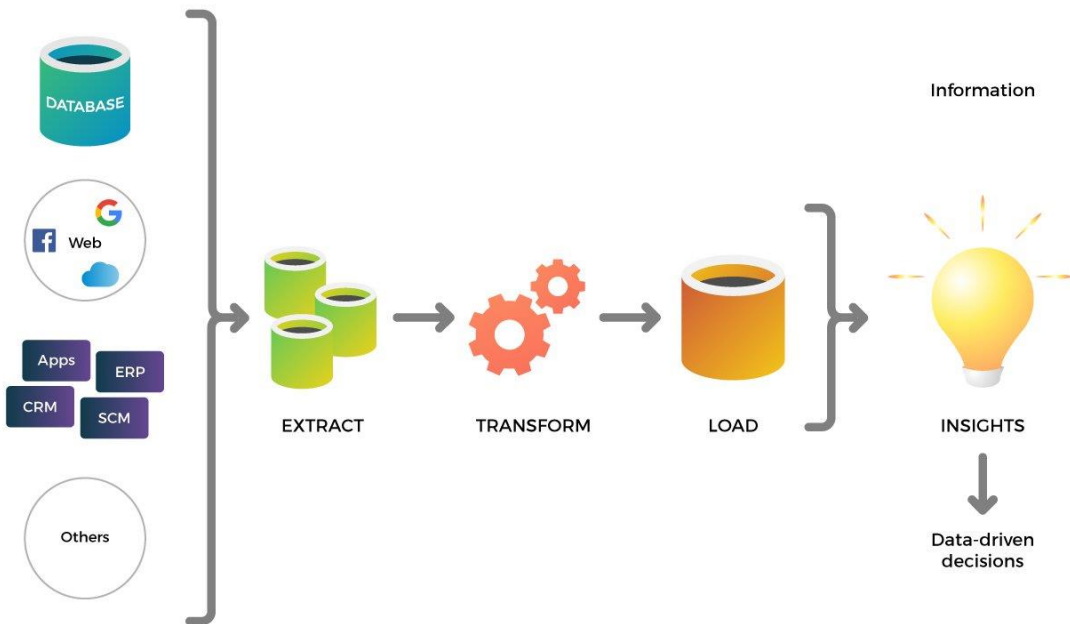


// ETL



ETL

ETI





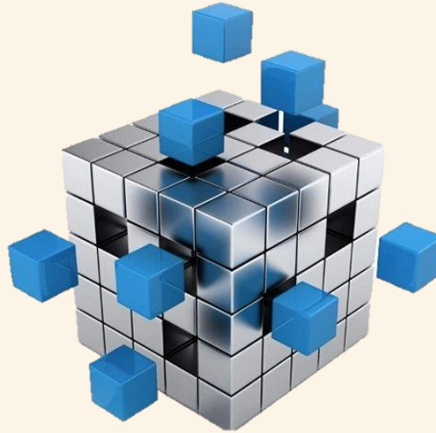
02

**MODELO DE
DADOS**

MODELO DE DADOS



É um modelo abstrato que organiza a descrição dos dados, a semântica dos dados e as restrições de consistência dos dados.

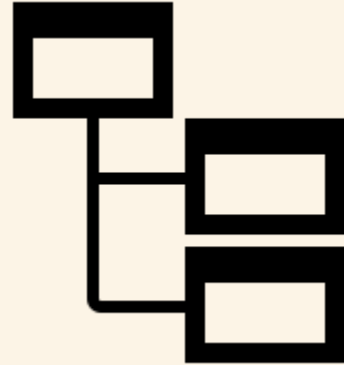


MODELAÇÃO DE DADOS



Os Modelos de Dados garantem

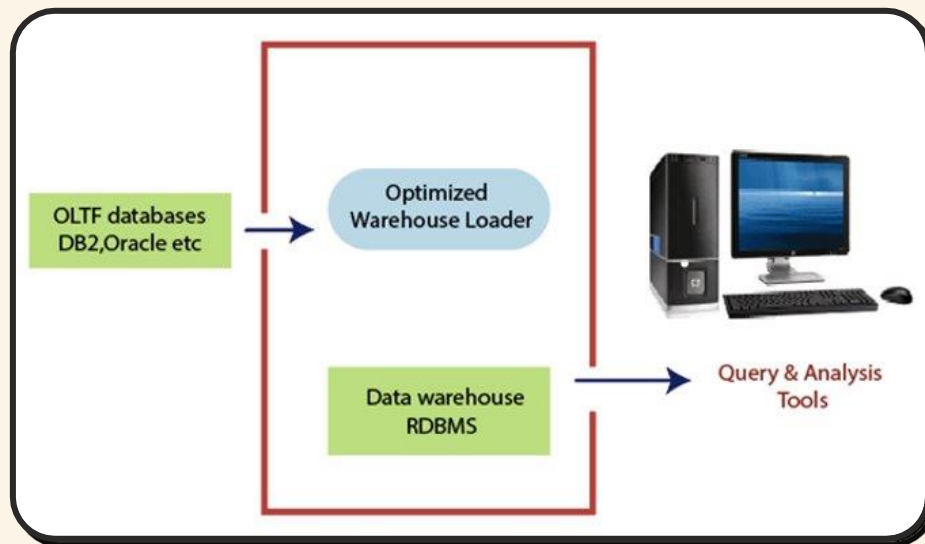
- a consistência nas convenções de nomenclatura,
- valores padrão,
- semântica,
- segurança,
- qualidade dos dados.



MODELO DE DADOS NA DATA WAREHOUSE

Os clientes do data Warehouse podem visualizar as relações entre os dados do Warehouse.

- Um esquema bem projetado equivale a uma estrutura de data Warehouse eficaz.



MODELOS DE DADOS NA DATA WAREHOUSE



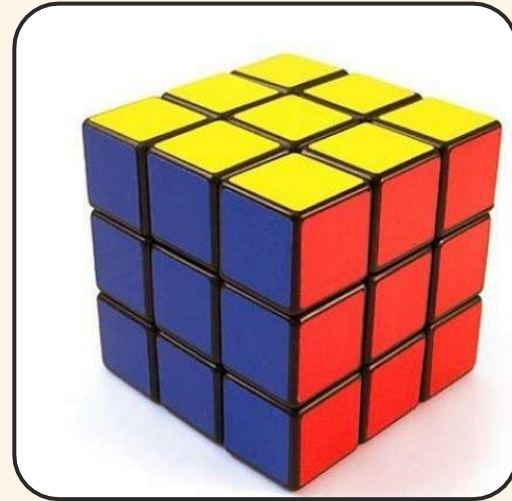
A data warehouse recorre a dois tipos de modelos de dados:



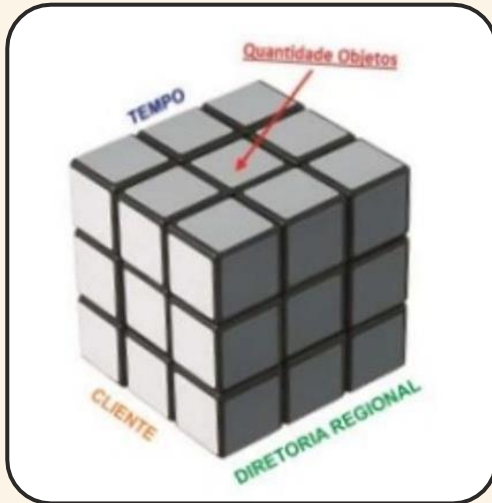
MODELO MULTIDIMENSIONAL



- É uma técnica de modelagem conceptual de negócios;
- Facilita a investigação, o resumo e a organização de dados para a análise de negócios;



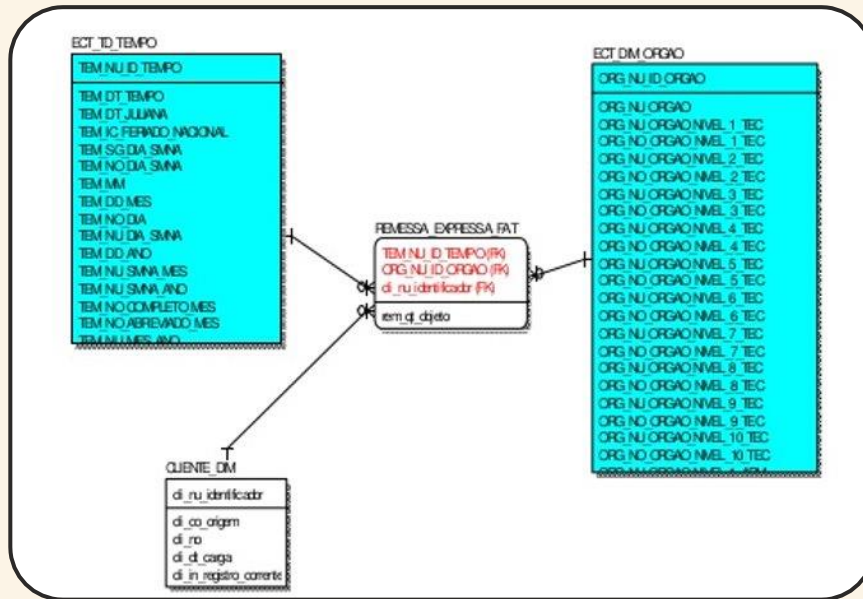
MODELO MULTIDIMENSIONAL



- Cruzamento de dimensões de análise para uma métrica
- Técnica de concepção e visualização de um modelo de dados

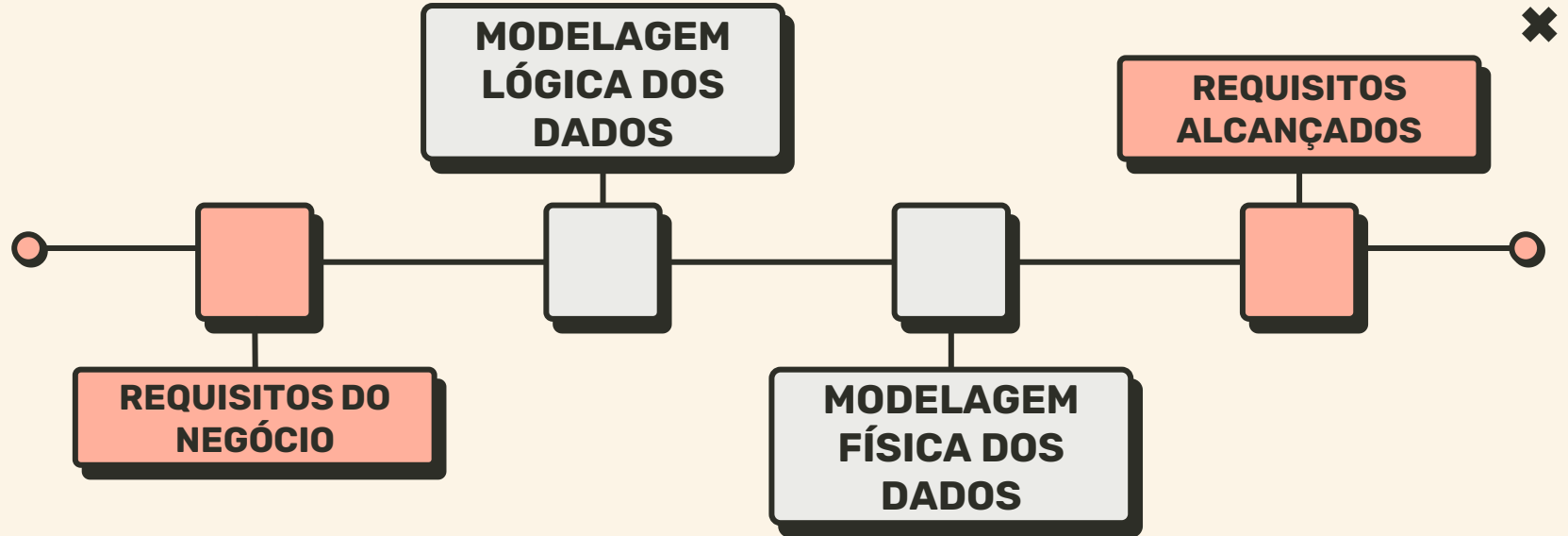


ELEMENTOS DE UM MODELO MULTIDIMENSIONAL ✕



- Dimensões
- Factos
- Métrica

CICLO DE VIDA DA MODELAGEM DE DADOS



TIPOS DE MODELOS DE DADOS



Existem 3 tipos de modelos de data warehouse

WAREHOUSE EMPRESARIAL

Armazena diversas
Informações de uma
organização/empresa

DATA MART

Projetado especialmente
para uma determinada
linha de negócios

WAREHOUSE VIRTUAL

Conjunto de percepções
sobre a base de dados
operacional

DATABASE VS DATA WAREHOUSE



OLTP	Usa	OLAP
Normalizadas	Tabelas	Desnormalizadas
Aplicações	Orientado para	Tema
Guardar dados	Objetivo	Analisar dados

The image features a central, large, cream-colored window with a black border and three small colored circles (red, yellow, green) in the top right corner. The window displays the number '03' in a large, bold, black font. Below the number, the text 'OLAP' is written in a bold, black font, set against a light red rectangular background. To the left of the main window, there are four smaller, cream-colored rectangular blocks stacked vertically. To the right, there are three more similar blocks, each with a black 'X' mark in its top right corner. The entire scene is set against a blue background with a repeating pattern of concentric, wavy lines.

03

OLAP



OLAP

- ONLINE ANALYTICAL PROCESSING

Sistema de armazenamento de dados agregados com capacidade para manipular e analisar um grande volume de dados sob múltiplas perspectivas.

VANTAGENS



FORNECE ÀS ORGANIZAÇÕES:



Método de aceder, visualizar e analisar os dados corporativos com alta flexibilidade e desempenho.

UTILIZADORES PODEM:



Rapidamente analisar inúmeros cenários



Gerar relatórios

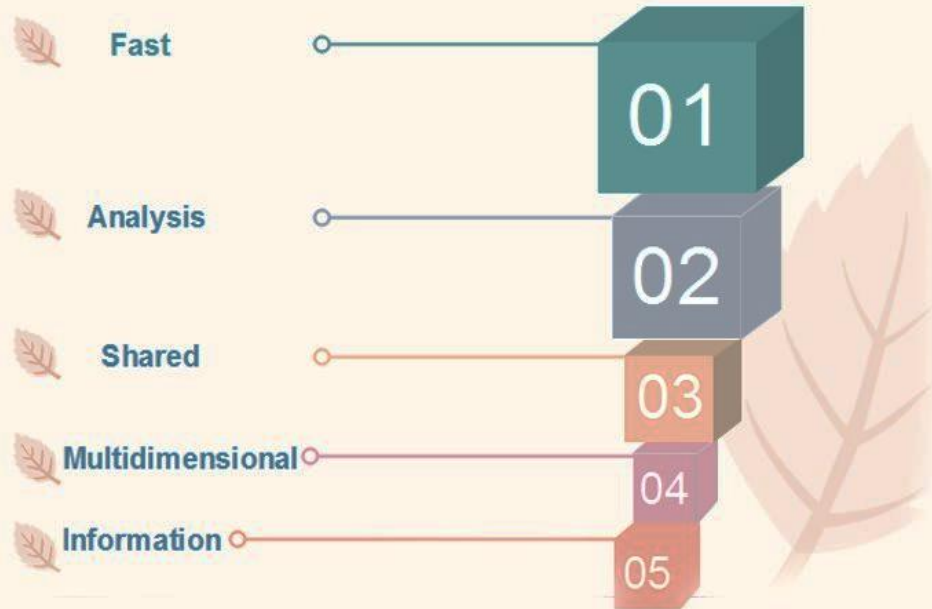


Descobrir tendências e factos relevantes

CARACTERÍSTICAS DO OLAP



Characteristics of OLAP



OLAP vs OLTP

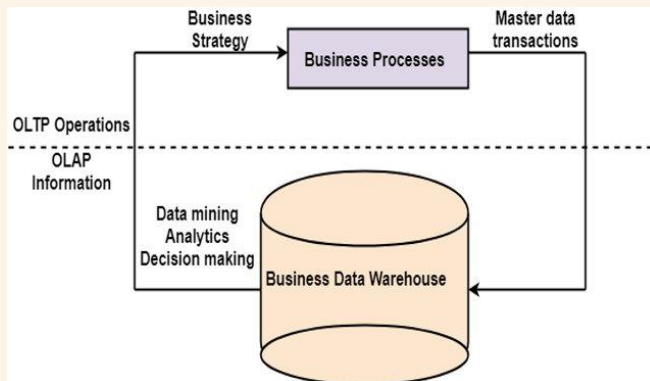
OLAP

Processamento Analítico On-line

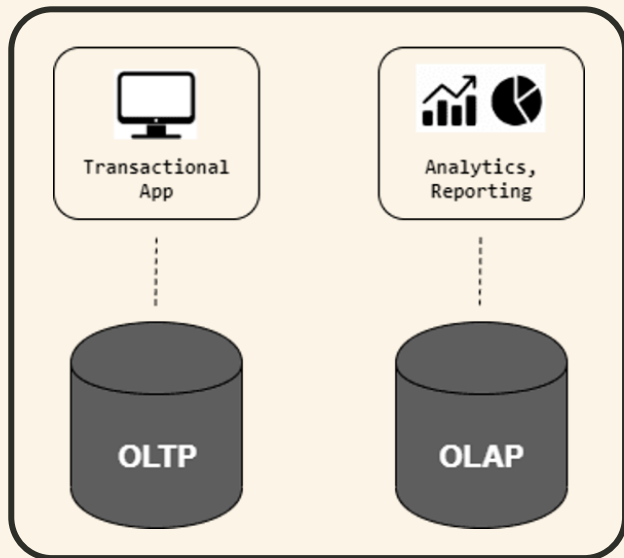
- É representado por um volume relativamente baixo de transações. As consultas são mais complexas e envolvem agregações.

OLTP

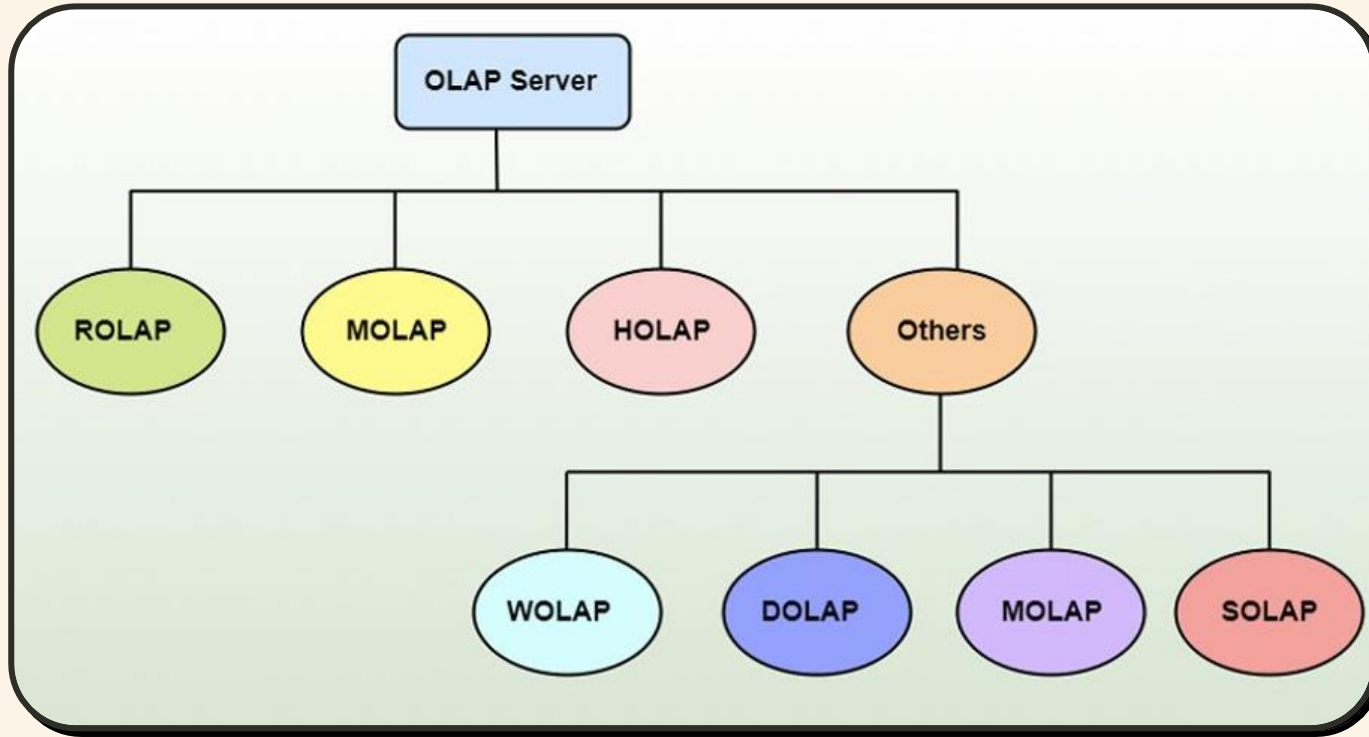
Processamento de transações on-line - É caracterizado por um grande número de transações curtas on-line (INSERT, UPDATE e DELETE).



Principais diferenças



MOLAP/ROLAP/HOLAP



Rolap

VANTAGENS

- ✗ Consegue suportar várias funcionalidades
- ✗ Não restringe o volume de armazenamento de dados.



DESVANTAGENS

- ✗ Baixo desempenho
- ✗ Limitado pelas funcionalidades SQL

Molap

VANTAGENS



Alto desempenho



Execução cálculos complexos



DESVANTAGENS



Investimentos altos.



Existem algumas restrições na limitação da quantidade dados a serem analisados

Holap

VANTAGENS



Alto desempenho



Não restringe o
volume de
armazenamento



DESVANTAGENS



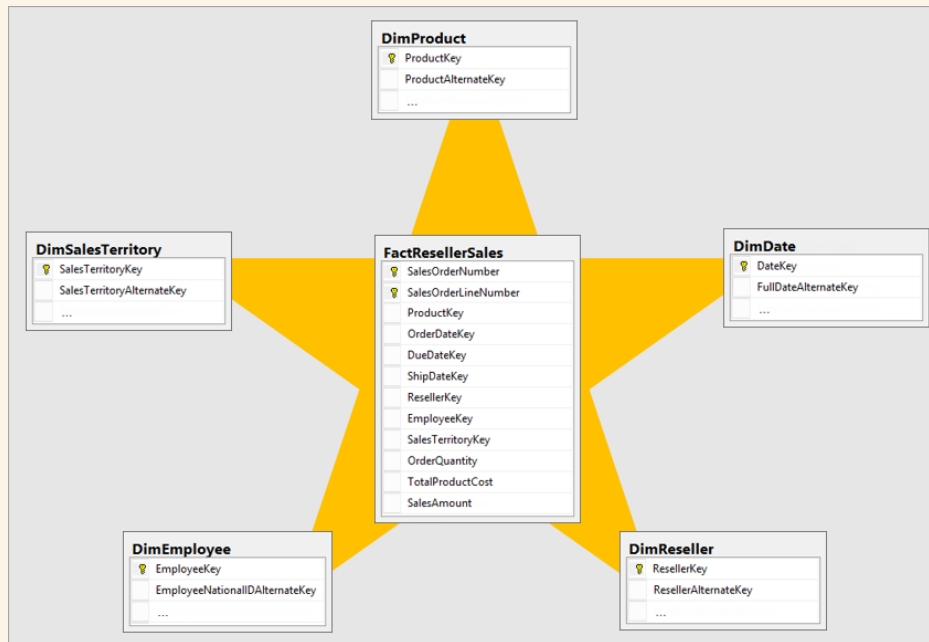
Maior custo

ESQUEMAS

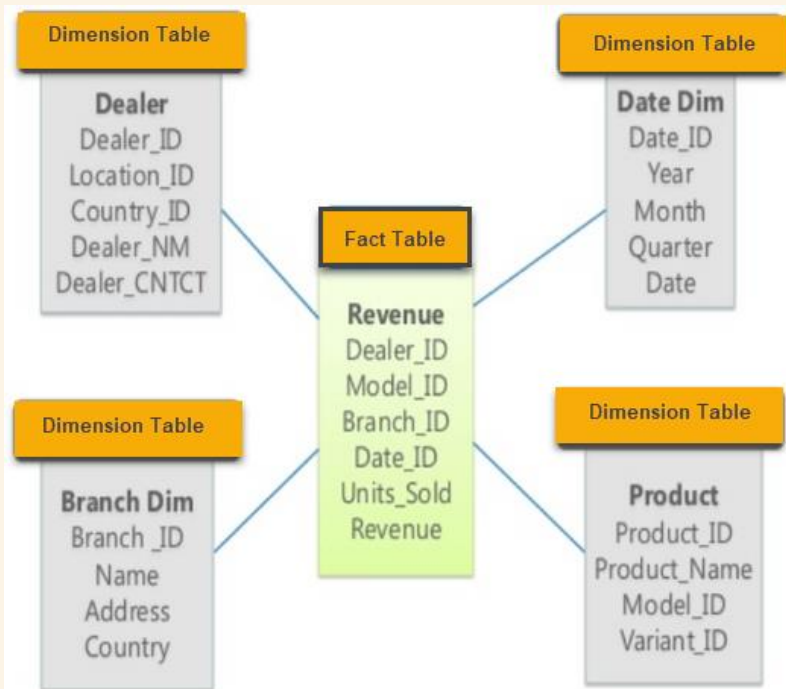




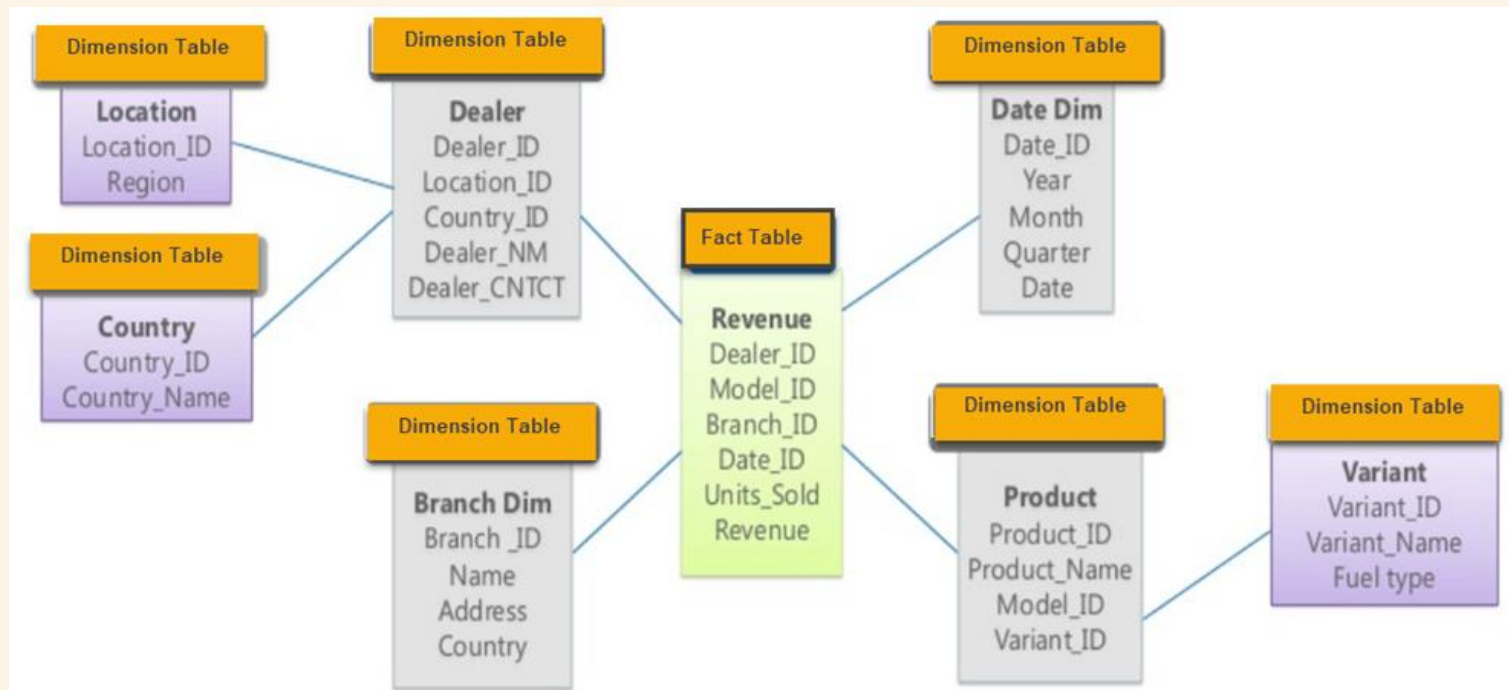
ESTRELA



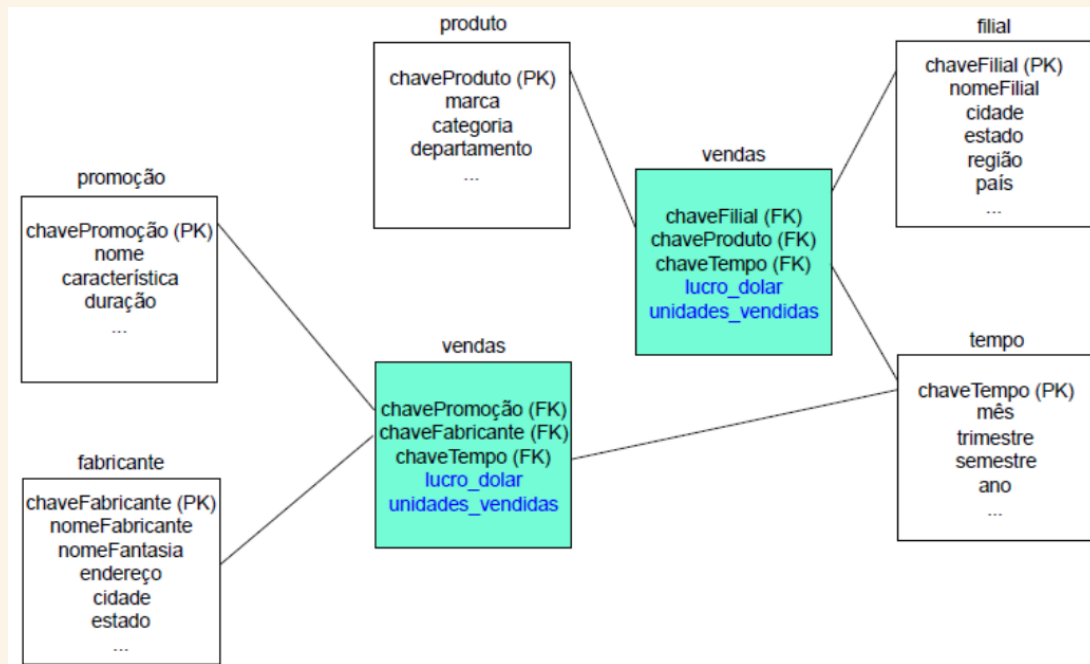
ESTRELA



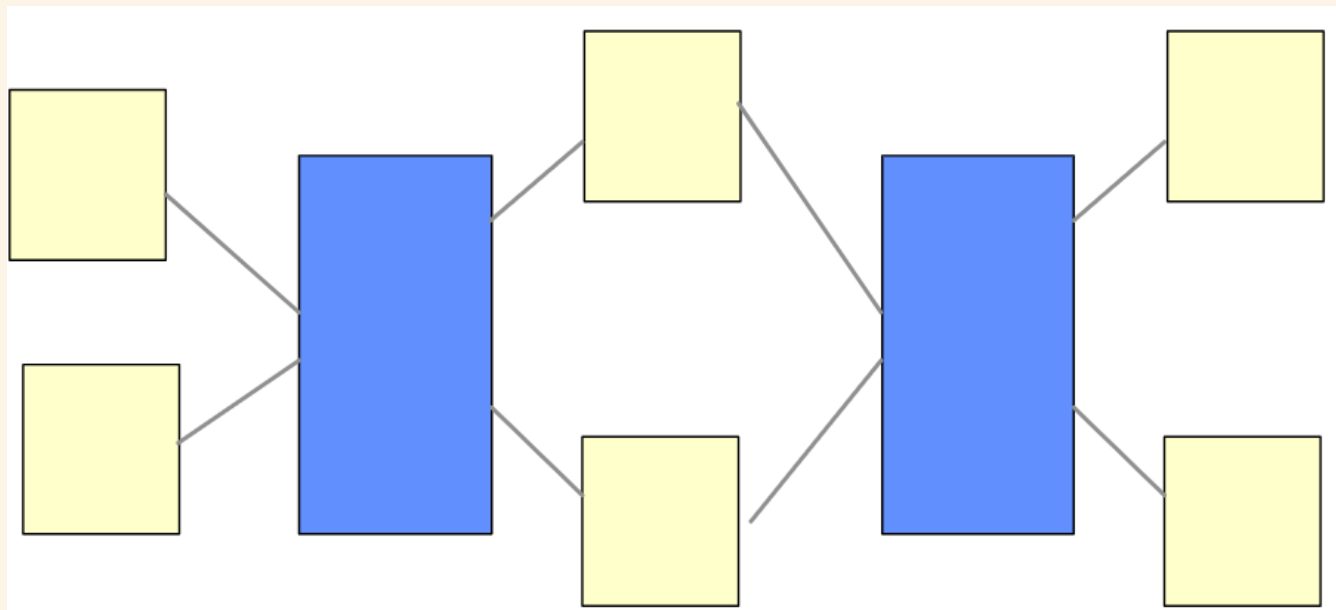
FLOCO DE NEVE



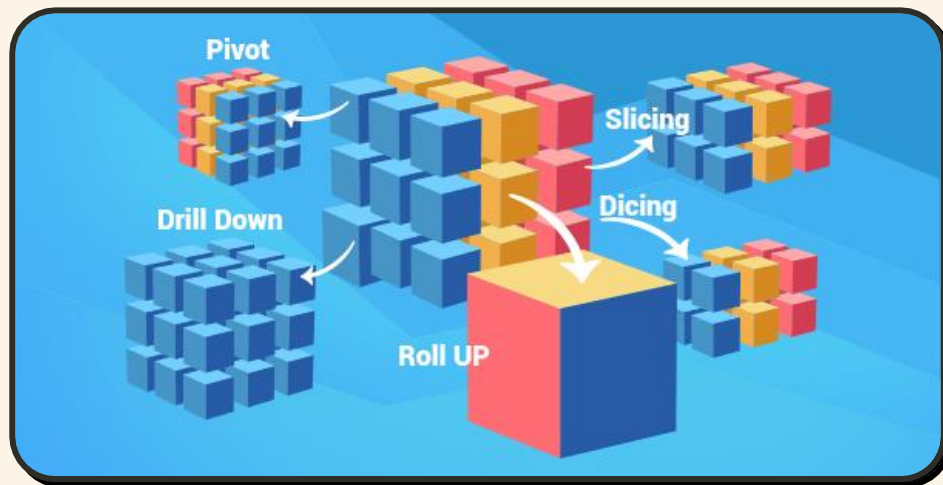
CONSTELAÇÃO DE FACTOS



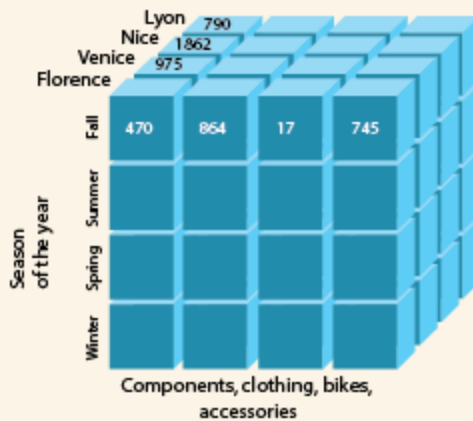
CONSTELAÇÃO DE FACTOS



OPERAÇÕES



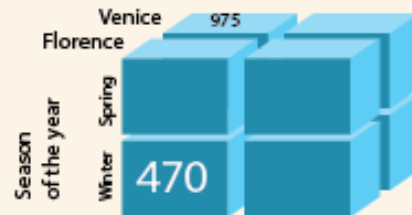
SLICE



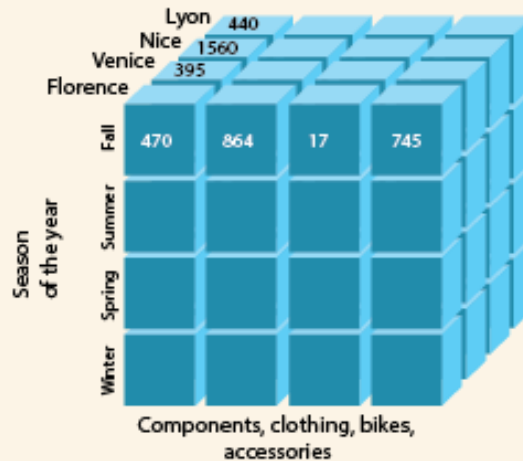
Slice
for time
="winter"



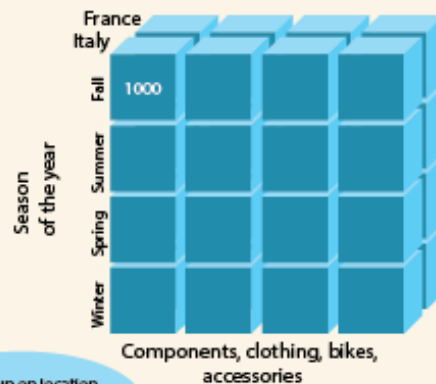
DICE



Dice for (location = "Venice" or "Florence")
and (season = "Winter" or "Spring") and
(item = "components" or "clothing")



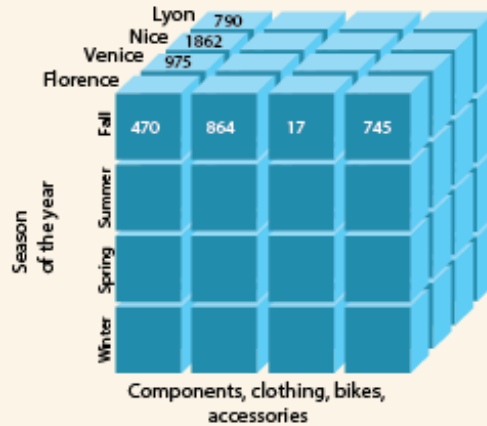
ROLL UP



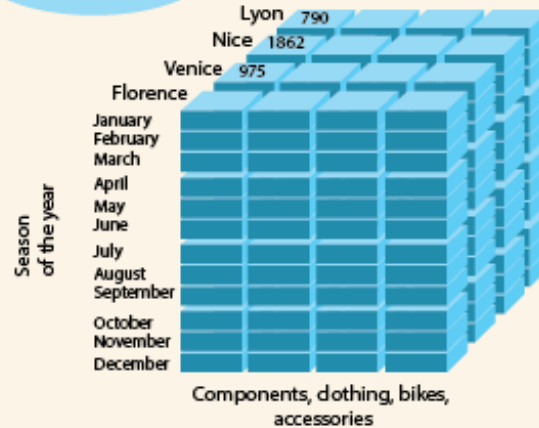
Drill-up on location
(from cities to
countries)



DRILL DOWN



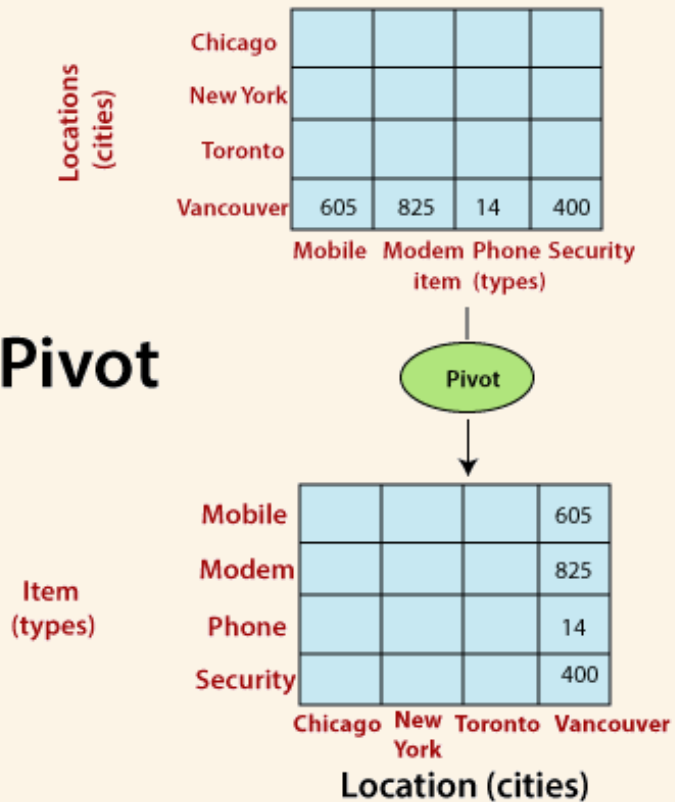
Drill down on time (from quarters to month)



PIVOT



Pivot



AS 14 OPERAÇÕES OLAP



DRILL-UP
DRILL-DOWN
SLICE
DICE
PIVOT
Scoping
Screening

Drill across
Drill through
Sort
Add measure
Drop measure
Union
Difference





04

MINERAÇÃO
DE DADOS

MINERAÇÃO DE DADOS - DEFINIÇÃO



- Método para descobrir padrões em enormes quantidades de informação.
- Mistura computação e estatísticas, e utiliza bastante machine learning e inteligência artificial.
- Dito isso, é uma buzzword.



BUZZWORDS

A buzzword is a word or phrase that is popular at a particular time or in a particular situation.

www.rakacreative.com

MINERAÇÃO DE DADOS - DEFINIÇÃO



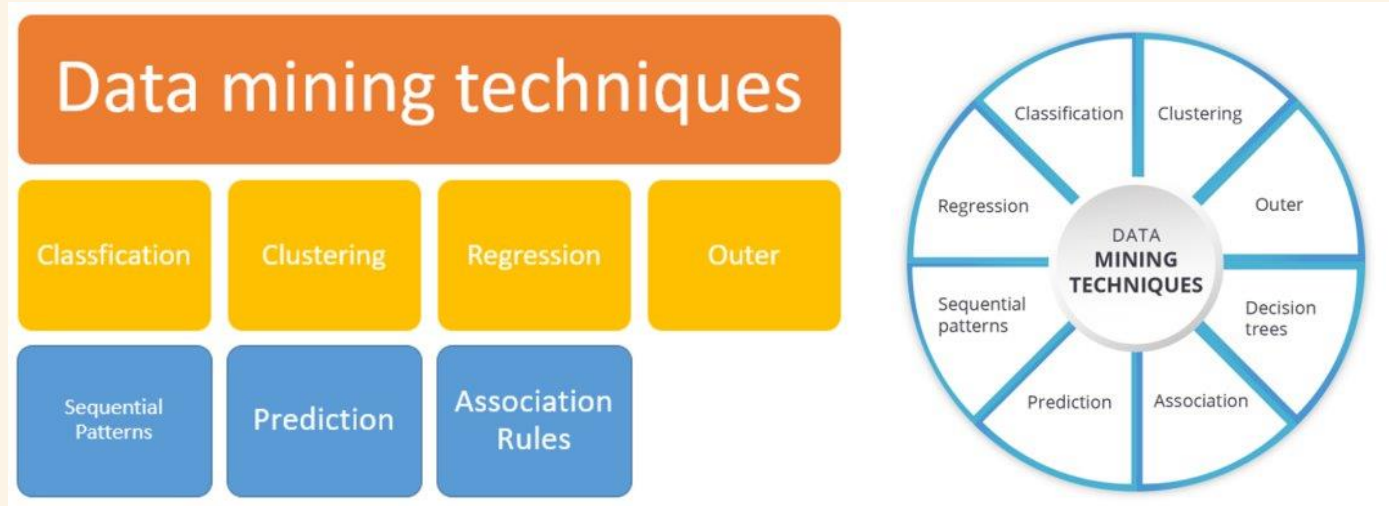
- Quase tudo o que envolva trabalhar com informação e AI/ML é chamado "data mining"
- Data mining refere-se apenas à extração de dados a partir de informação já existente.



TÉCNICAS



Informação descoberta varia bastante, mas há algumas técnicas em comum



TÉCNICAS

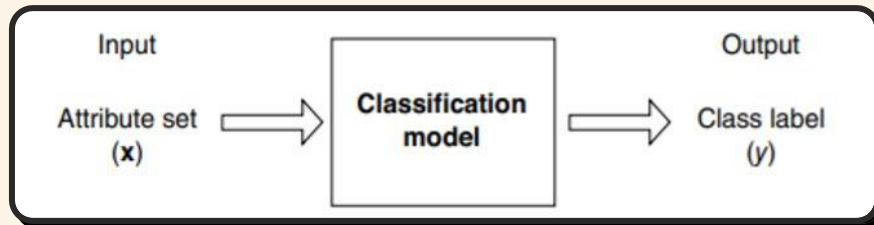


CLASSIFICATION

Colocar informação em categorias diferentes de acordo com certas características

Aplicações:

- Modelos descritivos
- Modelos preditivos

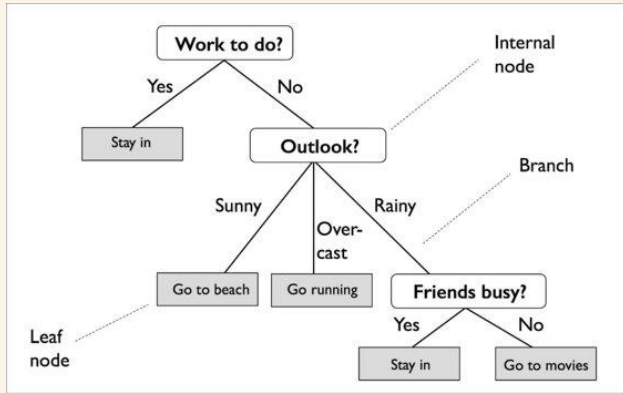


Exemplos:

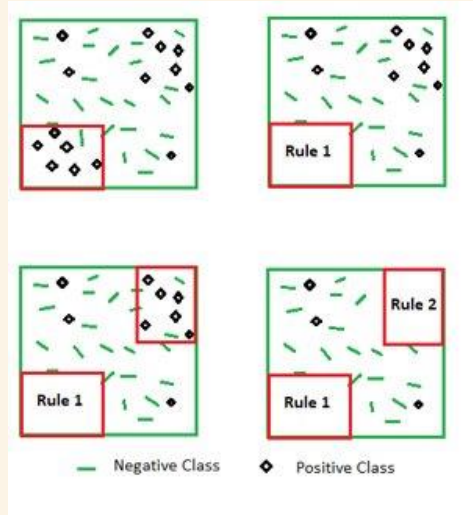
- Classificar email como "spam" ou "não-spam"
- Classificar células como malignas ou não-malignas



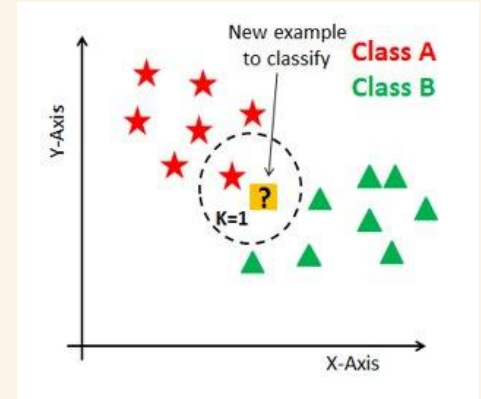
// CLASSIFICATION TECHNIQUES



DECISION TREES



RULE-BASED



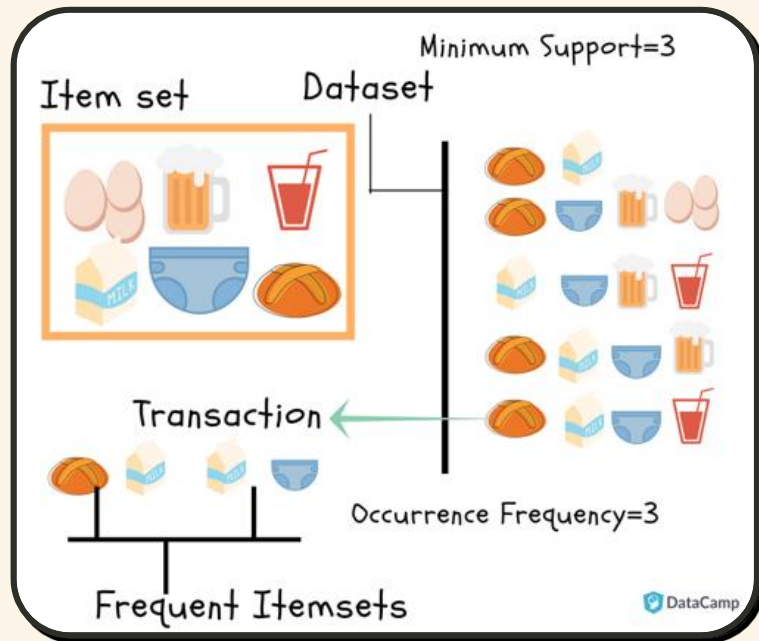
NEAREST-NEIGHBOR

TÉCNICAS



ASSOCIATION ANALYSIS

Descobrir ligações entre certas partes da informação



TÉCNICAS

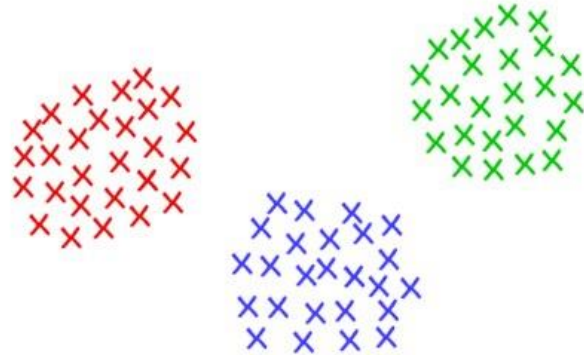


CLUSTER ANALYSIS

Agrupar informação em grupos de acordo com informação que tenham em comum

Usado para:

- Melhor categorizar partes da informação de modo conceptual;
- Modelos preditivos



TÉCNICAS

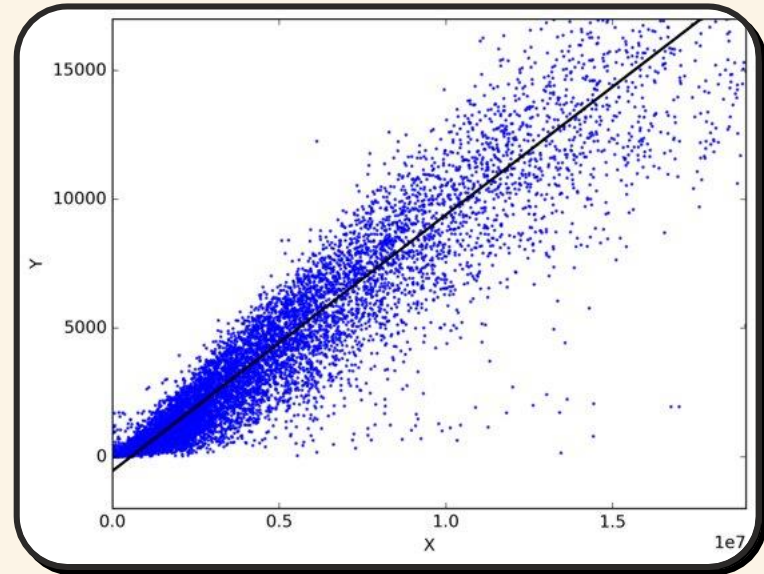


REGRESSION

Fazer previsões de tendências futuras com base em informação atual

Exemplos:

- Projetar valores de vendas para empresas
- Prever variações de ações no mercado



TÉCNICAS



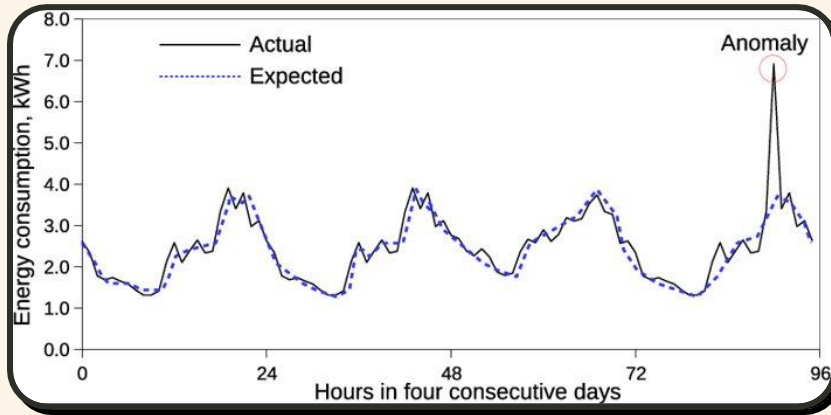
OUTER(ANOMALY) DETECTION



Conseguir descobrir e explicar/eliminar anomalias dentro da informação

Exemplos:

- Descobrir transações financeiras fraudulentas
- Detetar problemas de saúde em indivíduos

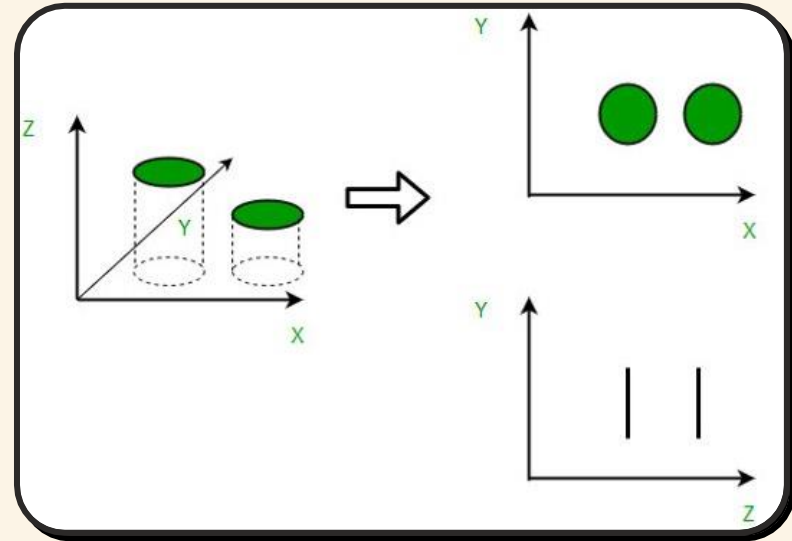


TÉCNICAS



DIMENSIONALITY REDUCTION

Simplificar problemas complexos diminuindo o número de variáveis a tratar



TÉCNICAS



OPTIMIZATION

Permite maximizar ou minimizar uma certa função

Exemplos:

- Descobrir a quantidade ideal de recursos para dispensar numa certa tarefa
- Descobrir como minimizar perdas de dinheiro

