

# Machine Learning & Chatbot

Conceitos Básicos Aula 1

Prof Alan

# Informações Gerais

- Terças, Quintas e Sextas feiras -> 3 aulas

07:15 as 10:00 / 19:10 as 22:00

Turmas : ADS4MA

LAB 2 de Informática da FECAF (CENTRO)

ADS4NA

LAB 3 de Informática da FECAF (CENTRO)

ADS4NB

LAB 4 de Informática da FECAF (PRINCIPAL)

## • EMENTA

- Conceitos fundamentais e básicos
- Estatística Básica
- Introdução a Ciência de Dados e ao Machine Learning Supervisionado
- Numpy (py) / Pandas (py) / Matplotlib (py) / Scikit Learn (py) e outras.
- Machine Learning Supervisionado (py) – Regressão
- Machine Learning Supervisionado (py) – Classificação
- Conceitos básicos de IA
- Conceitos de RBC
- Introdução ao CHATBOT
- Testes em LAB (CHATBOT)

# •BIBLIOGRAFIA

- **CRUZ, Leôncio Teixeira.** Assistentes Virtuais Inteligentes e Chatbots. São Paulo: Brasport, 2019.
- **VALDATI, Aline de Brittos.** Inteligência artificial - IA. Curitiba: Contentus, 2020.
- **BASSO, Douglas Eduardo.** Big Data. Curitiba: Contentus, 2020.
- **LUGER, George F.** Inteligência artificial. 6. ed. São Paulo: Pearson, 2013.
- 
- **MUNIZ, Antonio,** et al. Livro Jornada Phyton. São Paulo: Brasport, 2022.

# Tecnologias

## Inteligência Artificial

Se refere amplamente a qualquer comportamento semelhante ao do humano apresentado por uma máquina ou sistema.

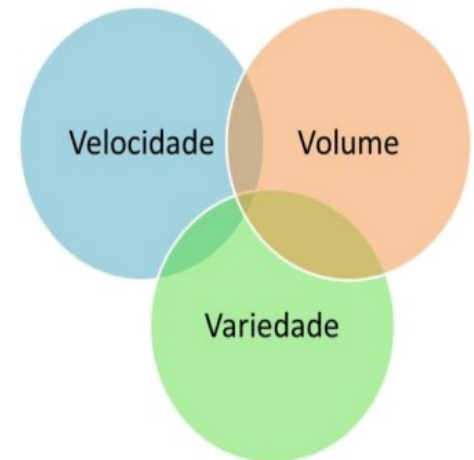
Na forma mais básica, os computadores são programados para “imitar” o comportamento humano.

Alguns “programas/modelos” têm **capacidade de aprendizado** com as suas experiências/exemplos (Aprendizado de Máquina)

## Big Data

Conjunto de dados extremamente amplos que, por isto, necessitam de ferramentas especiais para comportar o grande volume de dados

Dados estes processados em informações que possibilitam uma análise ampla e em tempo hábil.



# Big Data

## DADOS ESTRUTURADOS



bases de dados

## DADOS NÃO ESTRUTURADOS



- **Dados estruturados** – aqueles que possuem algum padrão ou formato que pode ser usado na sua leitura e extração dos dados. Dados de bancos de dados, arquivos texto (sejam csv, txt, XML, JSON,...).
- **Dados não estruturados** – não possuem um formato padronizado para leitura, podem ser arquivos Word, Páginas de Internet, Vídeos, áudios, entre outros.

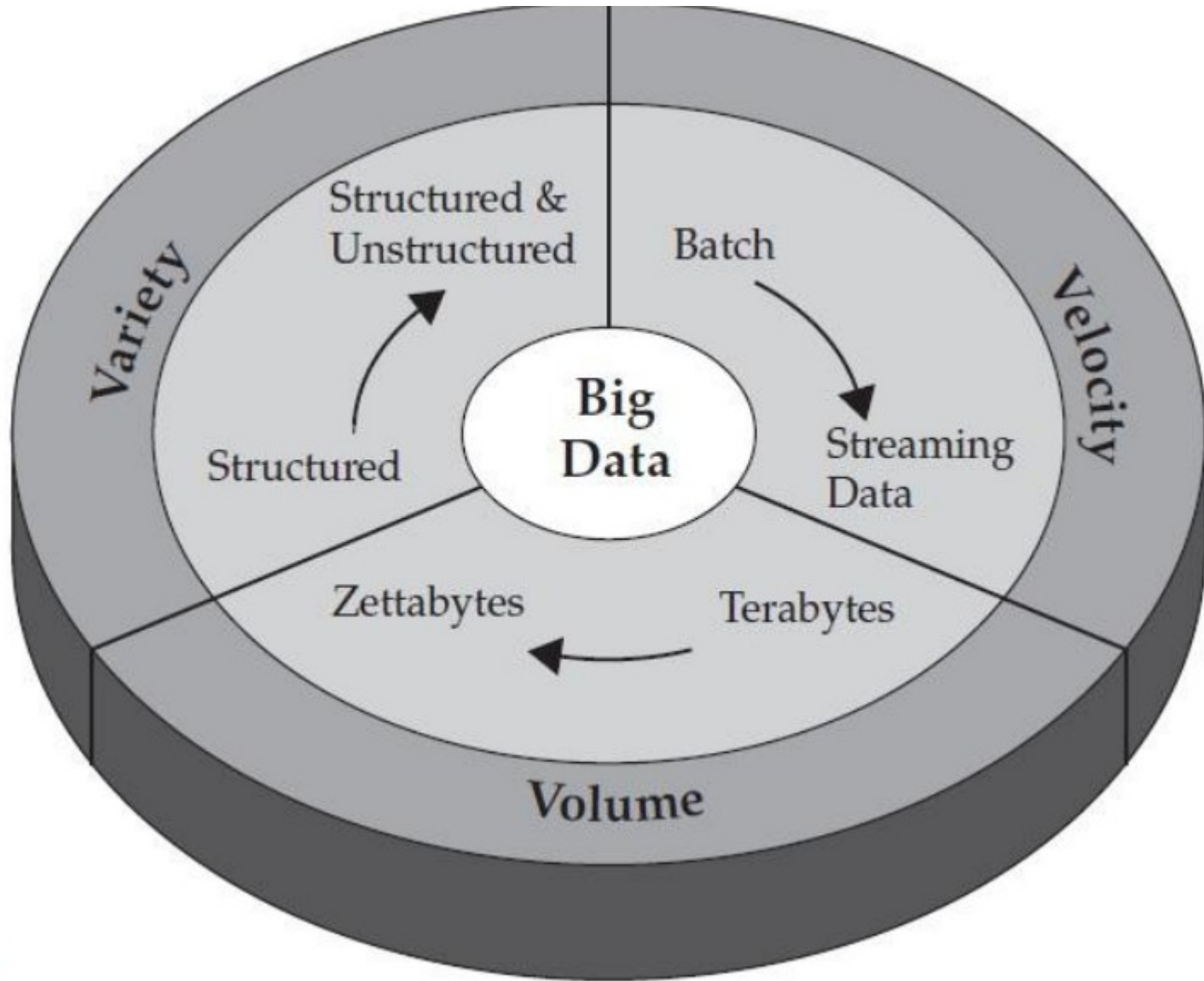
# Big Data - NoSQL

BDs orientados a documentos

Armazenam coleções de pares de chave-valor

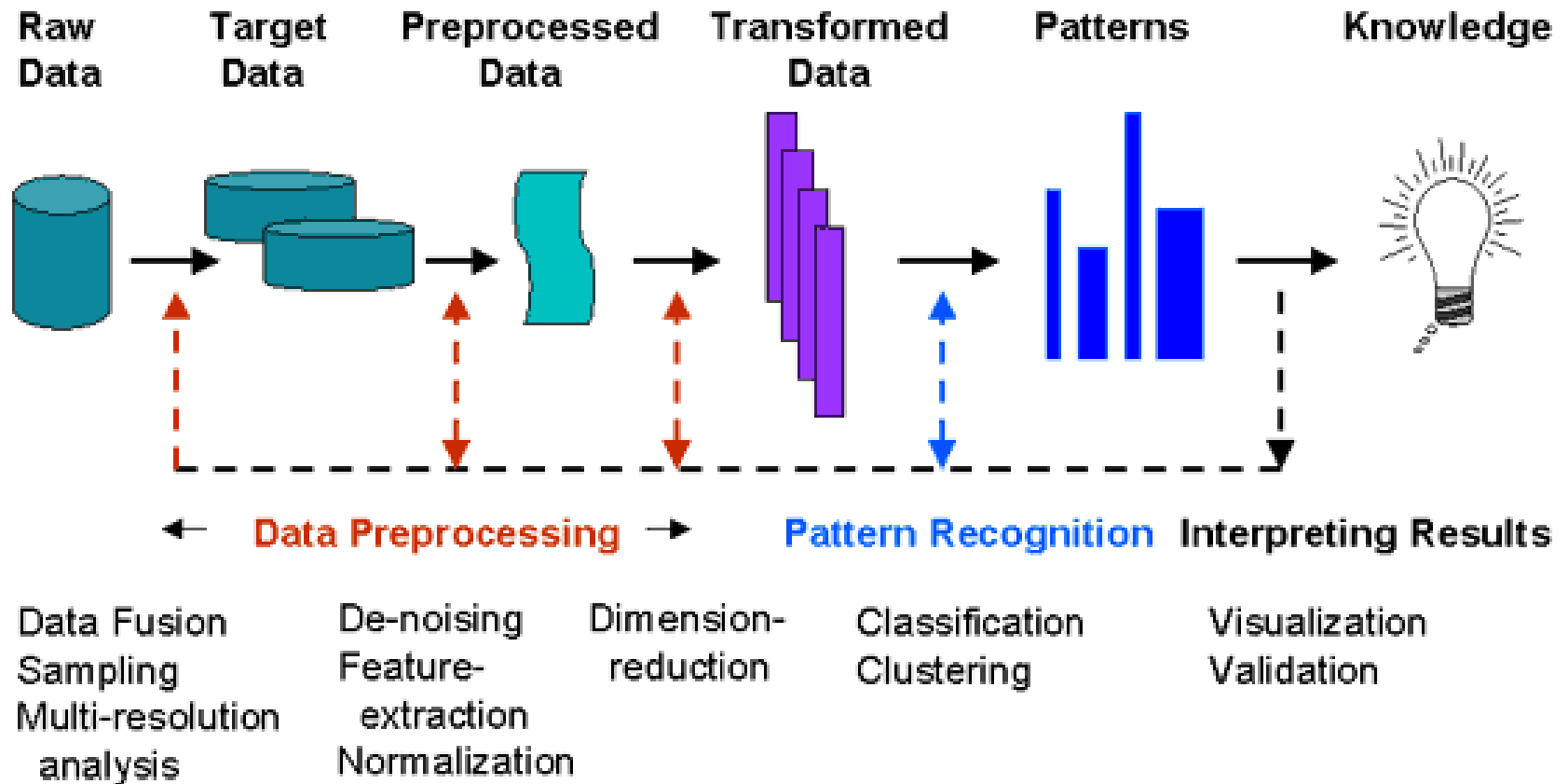
```
{  
  "id": 55,  
  "País": "Brasil",  
  "Região": "América do Sul",  
  "População": 201032714,  
  "PrincipaisCidades": [  
    {  
      "NomeCidade": "São Paulo",  
      "População": 1182876,  
    },  
    {  
      "NomeCidade": "Rio de Janeiro",  
      "População": 6323037,  
    }  
  ]  
}
```

Formato: JSON





# Mineração de Dados



**An iterative and interactive process**

# Mineração de Dados

## **Fraldas e cervejas**

- Homens casados, entre 25 e 30 anos compravam fraldas e/ou cervejas às sextas-feiras à tarde no caminho do trabalho para casa.
- Wal-Mart otimizou as gôndolas nos pontos de vendas, colocando as fraldas ao lado das cervejas

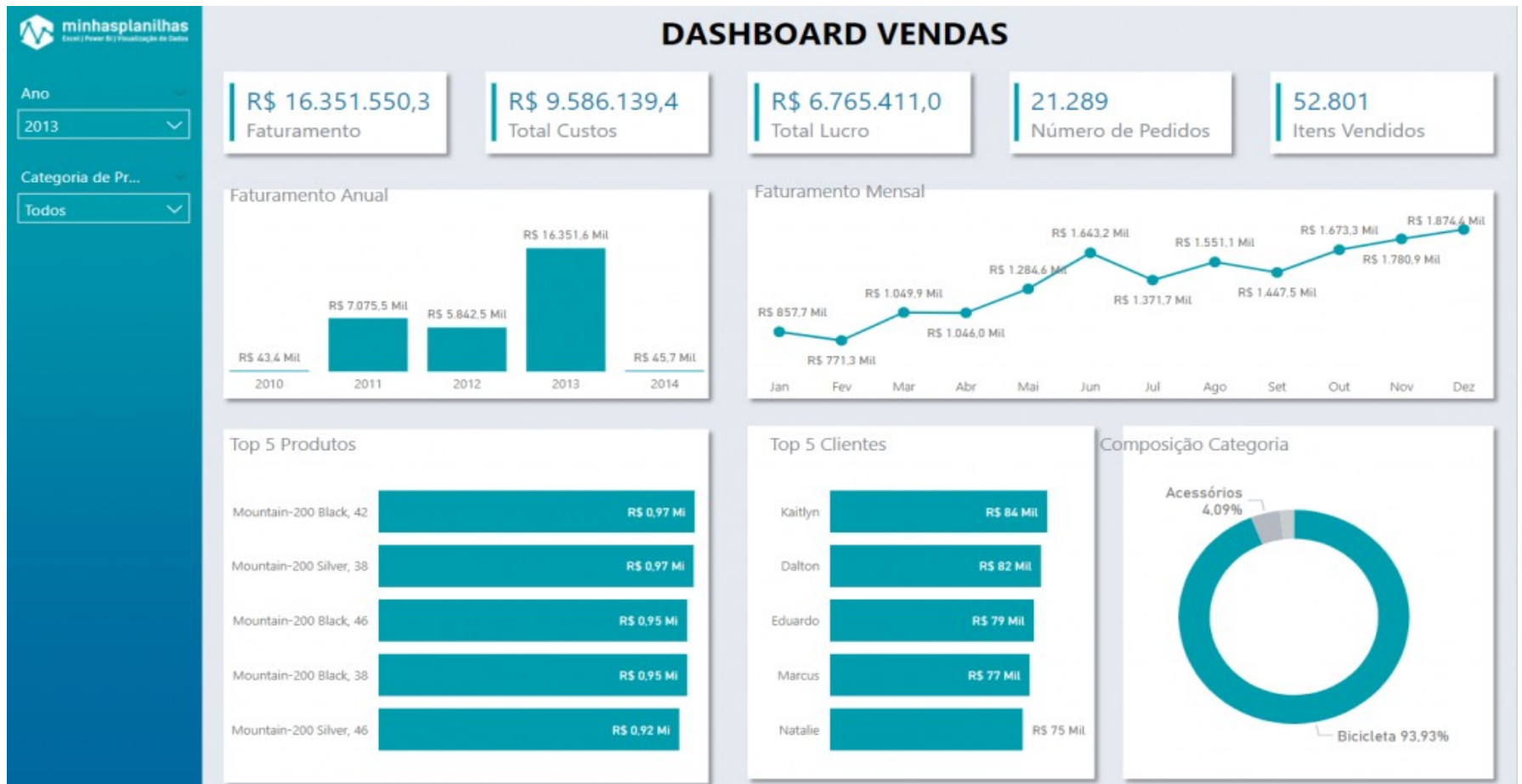
**Resultado: o consumo cresceu 30%**

## **Bank of America**

- Selecionou entre seus 36 milhões de clientes aqueles com menor risco de inadimplência (score)

**Resultado: em três anos o banco lucrou 30 milhões de dólares com a carteira de empréstimos.**

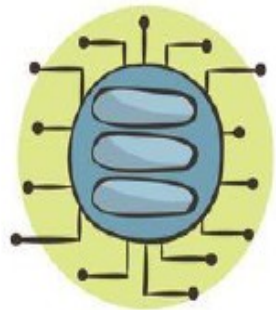
# Business Intelligence (BI)



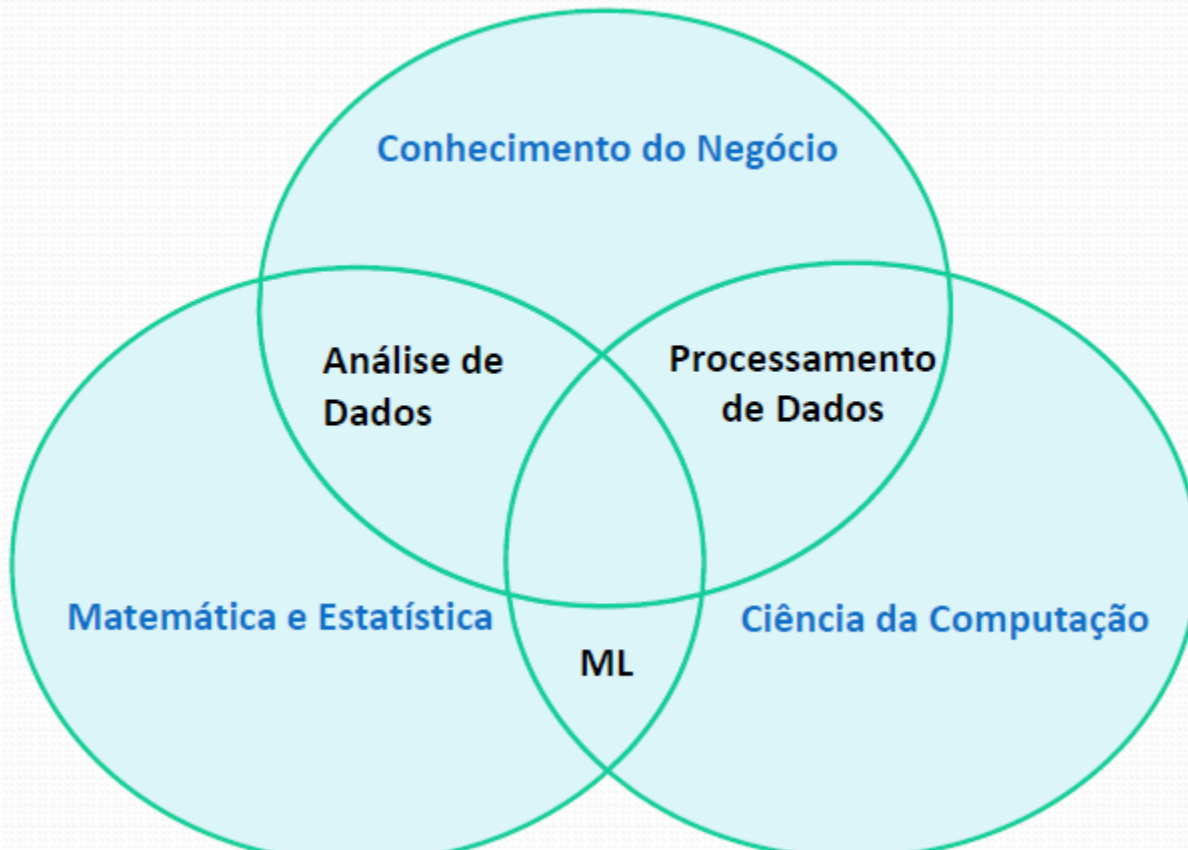


## O que é Ciência de Dados?

- Termo antigo mas que agora parece moderno (> 1960)
- Campo multidisciplinar que usa métodos científicos, processos, algoritmos e sistemas para extrair conhecimento e ideias de **dados convencionais e não convencionais**



## Quais são as áreas da Ciência de Dados?



# Data Science (MLS)

## Aplicação: Classificação de dígitos manuscritos (MNIST)



The MNIST Database

(Yann LeCun *et. al*)

Melhor desempenho

convnet, erro 0.23%

(Ciresan et al. CVPR 2012)

<http://yann.lecun.com/exdb/mnist/>

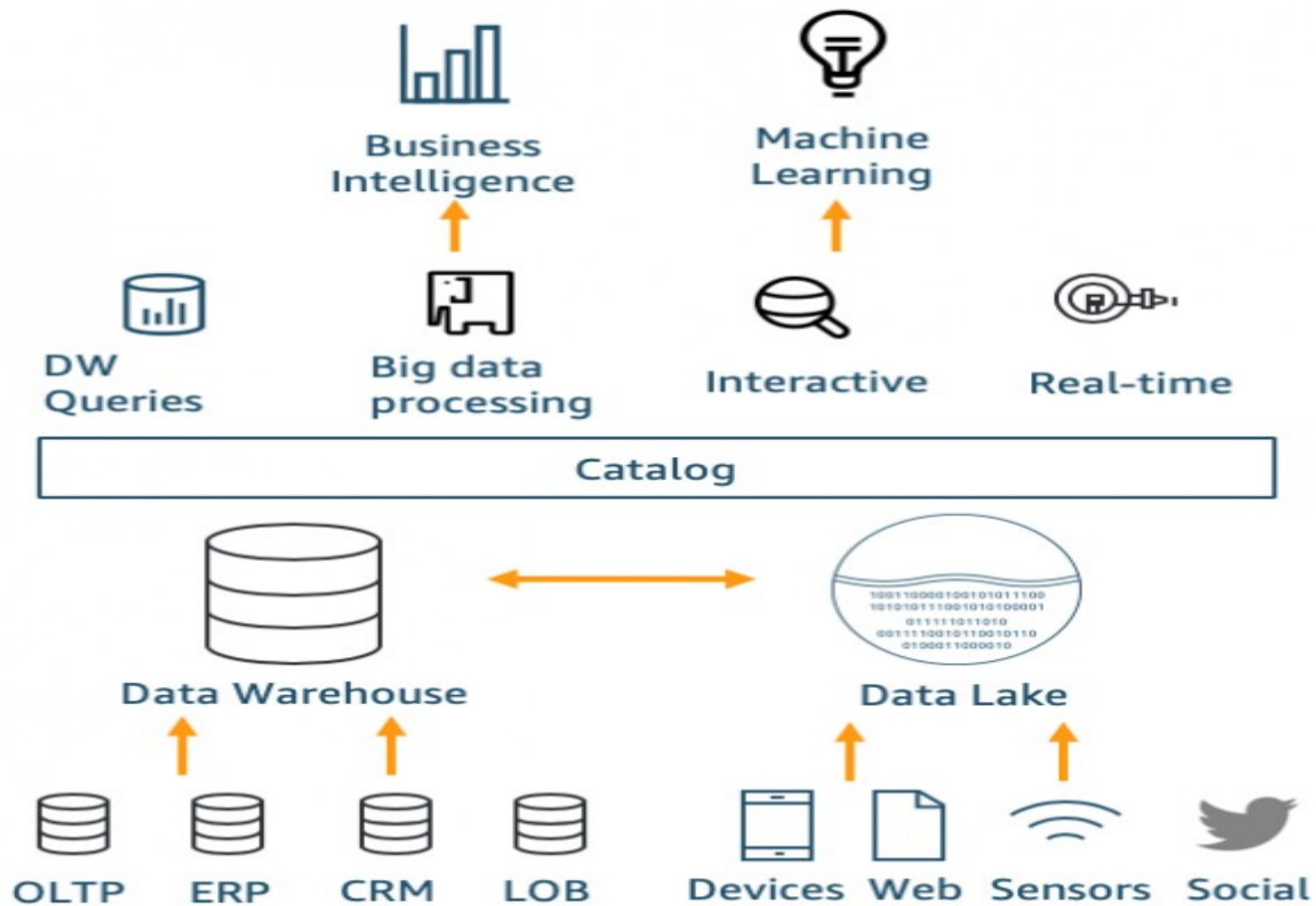


# O problema da Previsão



- O comerciante quer **prever** de acordo com fatos históricos, quantas cervejas ele venderá em uma janela de tempo
  - Tendo em mãos – número de compras por período
  - **Série temporal**



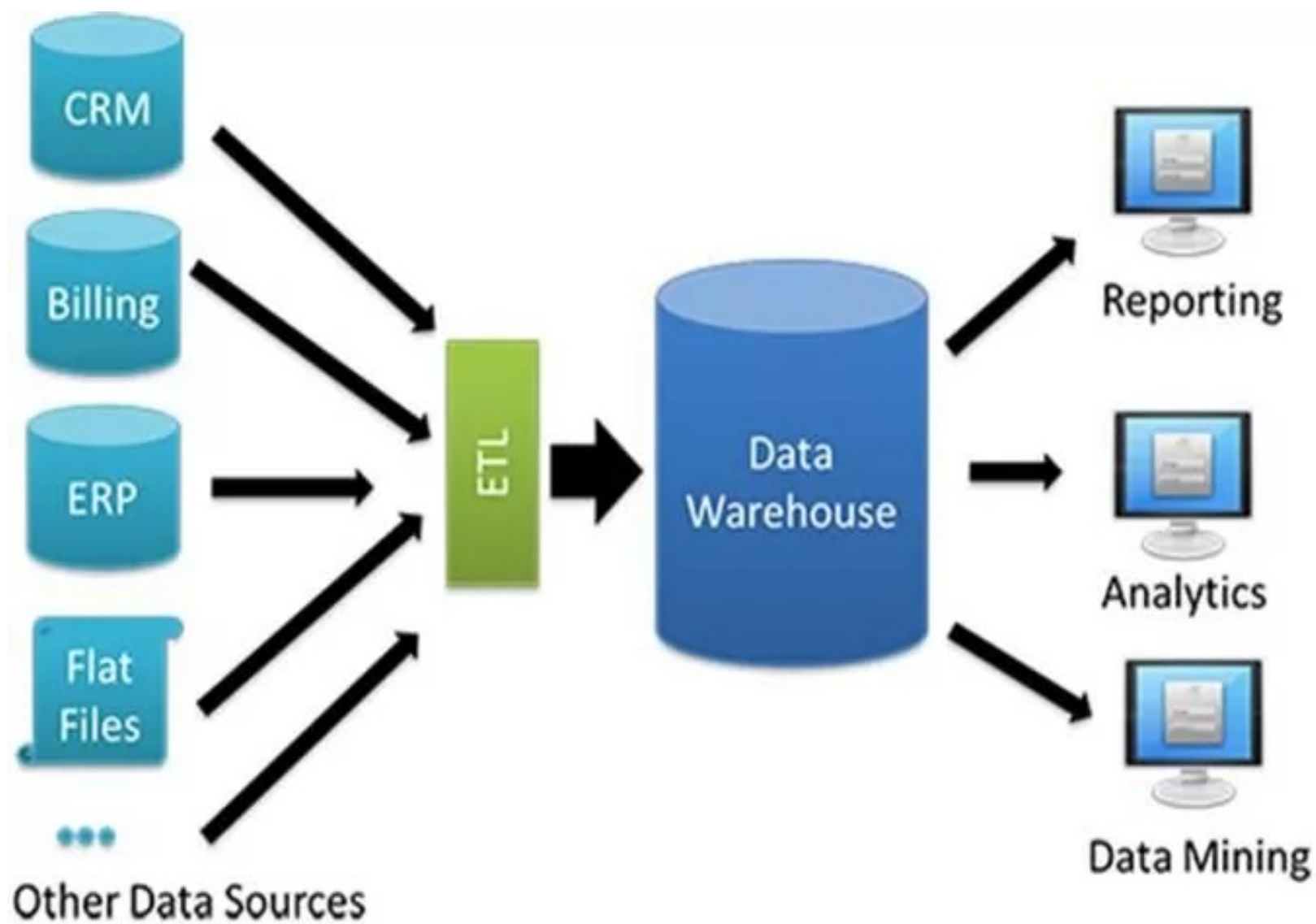


# Data Warehouse (Armazém de Dados)

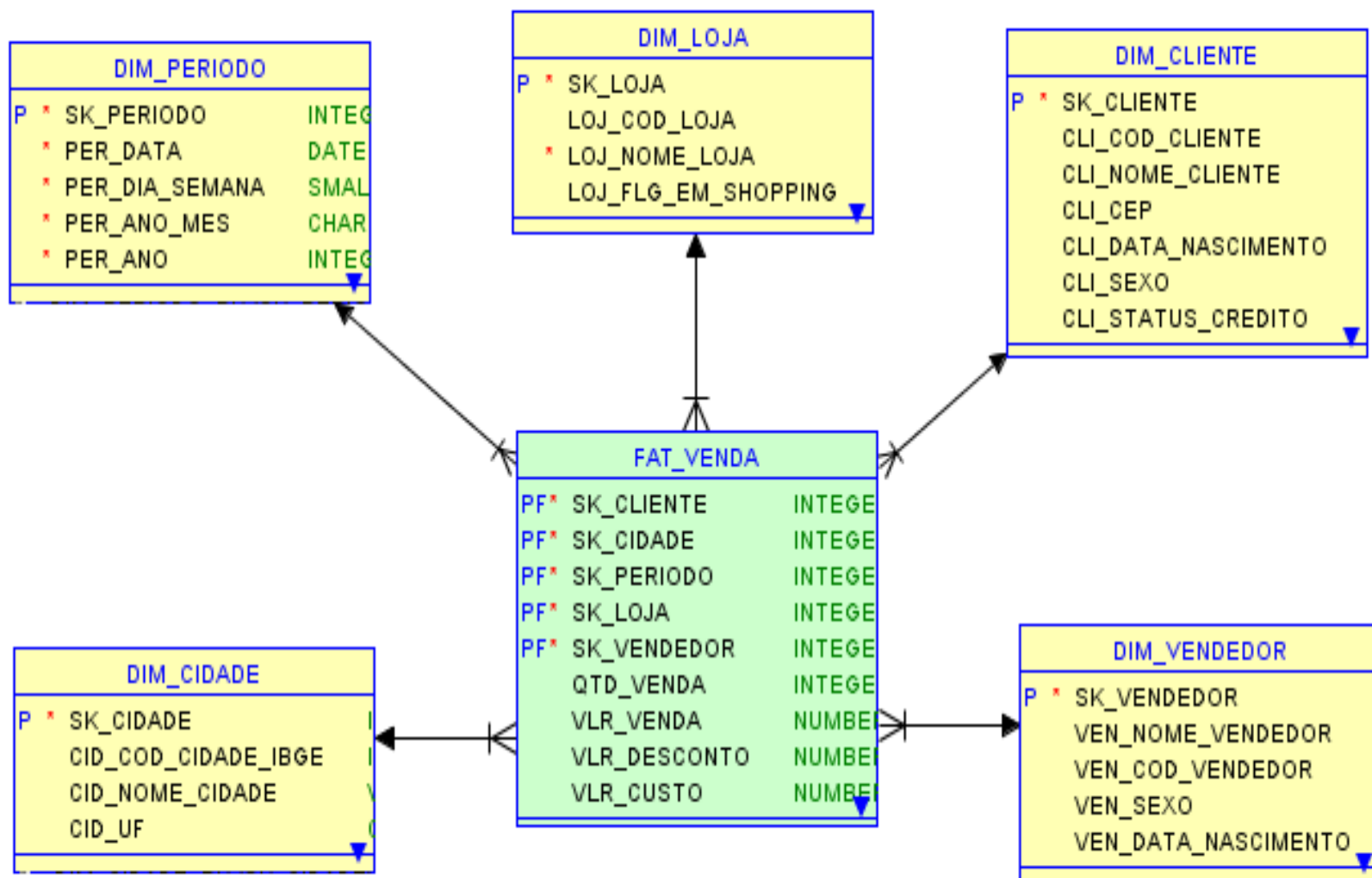
- Banco de dados voltado para o suporte à tomada de decisão
- Derivado de vários bancos/fontes de dados operacionais
- Na maioria das vezes é histórico
- Usa dados estruturados
- Usado como base para executar análises de informação e extração de conhecimento

## **Objetivos:**

- Análise de informações
- Monitorar e comparar situações atuais com passadas
- Estimar situações futuras



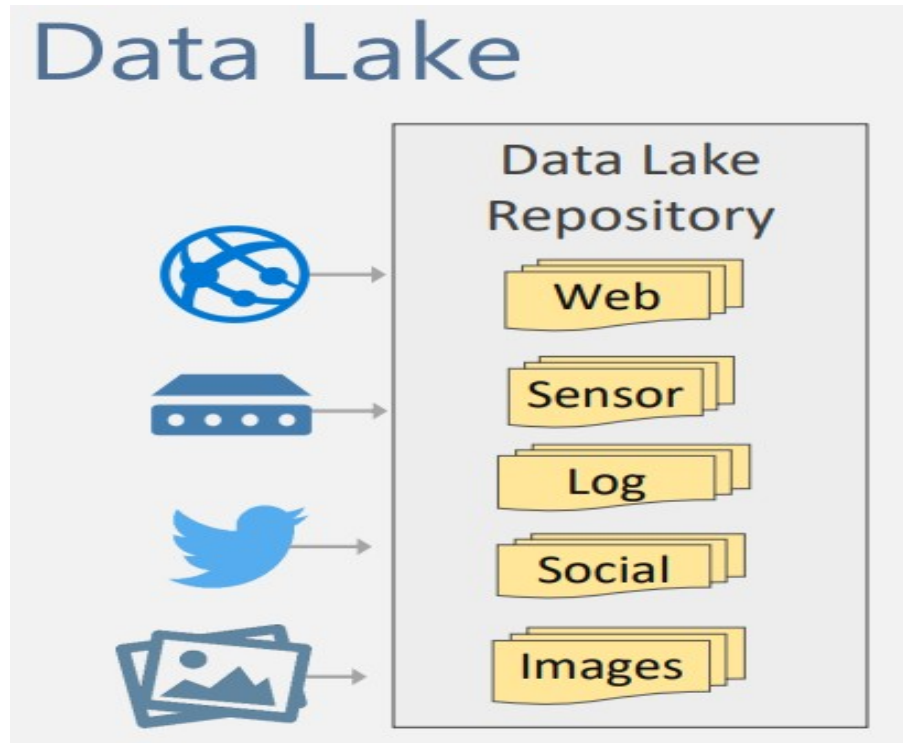
# Data Warehouse (Modelagem dimensional)



# Data Lake (Lago de Dados)

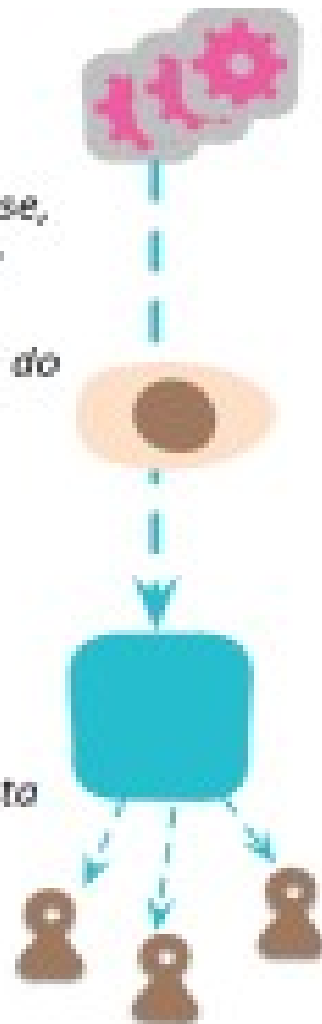
- É um repositório que armazena grandes quantidades de dados, vindos de diferentes fontes, no formato nativo (Sem modelagem).
- Considerada a fonte de dados do BIG DATA
- Armazena diversos tipos de dados (estruturados e/ou não estruturados)
  - lot
  - Tweets
  - Email
  - Dados estruturados
  - ...

# Data Lake (Lago de Dados)



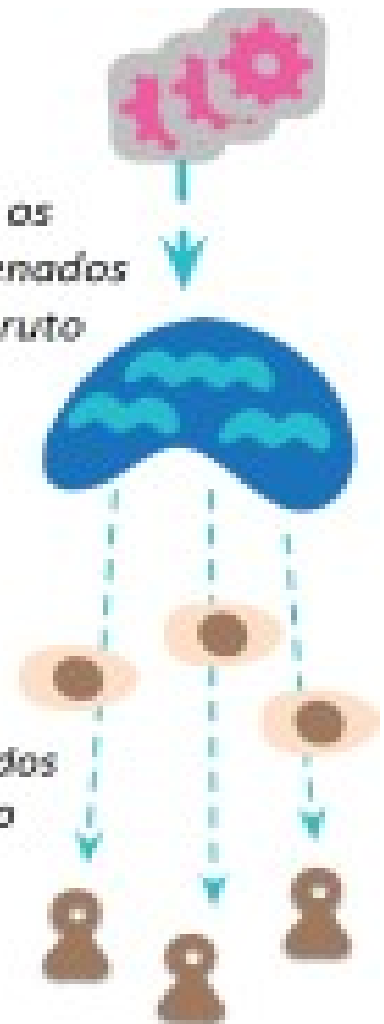
# Data Warehouse x Data Lake

*Com o Data Warehouse, os dados são limpos e organizados em um único esquema, antes do armazenamento*



*A análise é feita consultando diretamente no Data Warehouse*

*Com o Data Lake, os dados são armazenados em seu formato bruto*



*Os dados são selecionados e organizados de acordo com a necessidade*