



# Unidade I - Introdução à Estatística

## Estatística na Informática

## FUNDAMENTOS

# Duas Ciências, Um Objetivo Comum

## Informática

A parte que fornece as **ferramentas** — hardware e software — para coletar, armazenar, processar e gerenciar grandes volumes de dados com eficiência e velocidade.

## Estatística

A ciência que oferece os **métodos e técnicas** para analisar, interpretar, extrair conclusões significativas e tomar decisões fundamentadas baseadas nesses dados.



# O Poder da União

A convergência da Estatística e da Informática criou um campo de conhecimento extremamente poderoso, conhecido como **Ciência de Dados**, **Analytics** ou **Mineração de Dados (Data Mining)**.

## Informática Sozinha

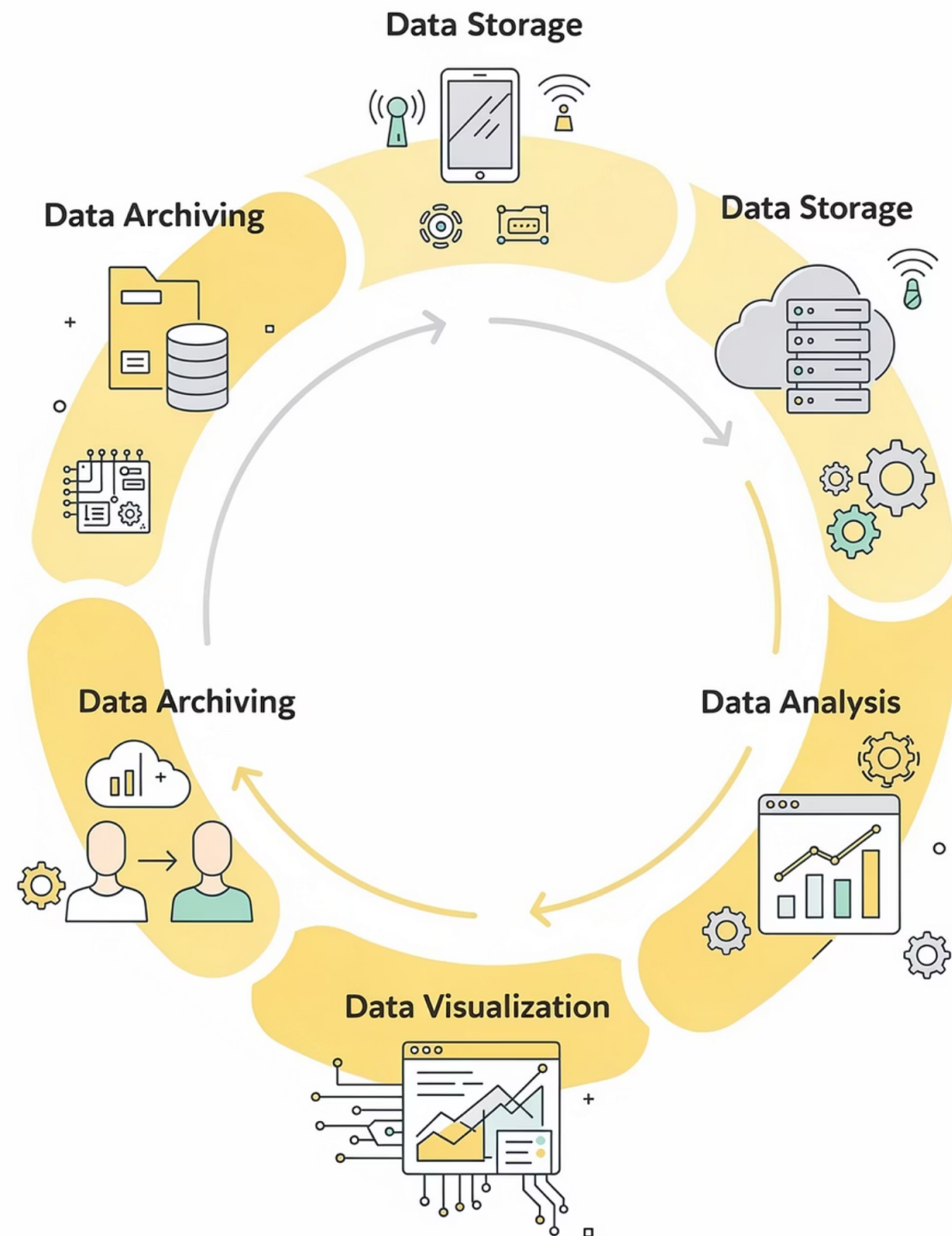
Consegue armazenar terabytes de dados, mas não sabe o que fazer com eles ou como extrair valor significativo.

## Estatística Sozinha

Tem dificuldade de processar volumes gigantescos de informação manualmente, limitando sua aplicação prática.

## Juntas

Se completam perfeitamente, criando soluções poderosas que transformam dados brutos em insights acionáveis.



## CICLO DE VIDA

# Como Elas se Relacionam na Prática

A relação entre Estatística e Informática acontece em todas as etapas do ciclo de vida de um sistema computacional que lida com dados. Vamos explorar cada uma dessas etapas fundamentais.



# Etapa 1: Coleta de Dados

A Informática cria os **sistemas** através dos quais os dados são gerados e coletados de forma contínua e automatizada. Essa é a porta de entrada para todo o processo analítico.

## Exemplos Práticos:

- Logs de servidor registrando cada acesso
- Cliques de usuário em aplicações web
- Formulários online capturando informações
- Sensores IoT gerando dados em tempo real
- APIs integrando diferentes fontes de dados



# Etapa 2: Armazenamento e Gerenciamento

A Informática desenvolve os **bancos de dados** (SQL, NoSQL) e **data warehouses** que guardam esses dados de forma estruturada, segura e acessível. Conceitos de amostragem estatística podem ser aplicados para otimizar consultas em volumes massivos.



## Bancos de Dados Relacionais

MySQL, PostgreSQL e Oracle para dados estruturados e transações complexas com garantias ACID.



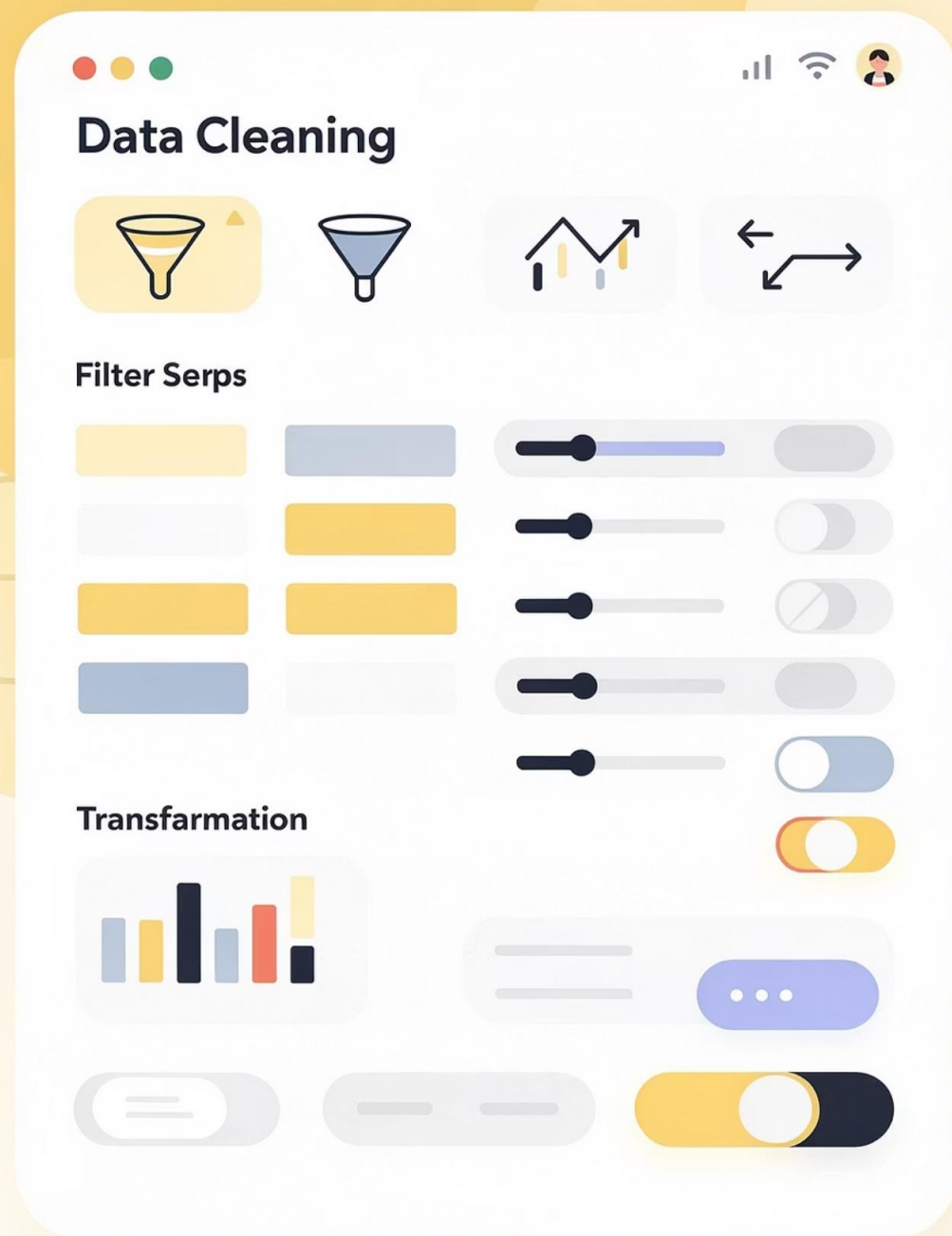
## Bancos NoSQL

MongoDB, Cassandra e Redis para dados não estruturados, escalabilidade horizontal e flexibilidade.



## Data Warehouses

Amazon Redshift, Google BigQuery para análises complexas e consultas otimizadas em grandes volumes.



## Etapa 3: Processamento e Limpeza

Antes da análise, os dados precisam ser tratados e preparados. A Informática fornece as **ferramentas** (bibliotecas como Pandas em Python, ou softwares como R) para a Estatística aplicar suas técnicas especializadas.

1

### Detecção de Valores Ausentes

Identificação e tratamento de dados faltantes usando métodos de imputação ou exclusão estratégica.

2

### Tratamento de Outliers

Análise de valores discrepantes para determinar se são erros ou observações legítimas.

3

### Correção de Inconsistências

Padronização de formatos, validação de tipos e resolução de duplicações nos dados.

## Etapa 4: Análise Exploratória

Aqui a união é total. Usamos ferramentas computacionais (linguagens de programação, softwares de visualização) para aplicar métodos estatísticos e explorar os dados em profundidade.

### Visualização de Dados

Criamos gráficos dinâmicos e interativos — histogramas, boxplots, gráficos de dispersão — para entender a distribuição e o relacionamento entre as variáveis. A Informática torna essa visualização rica e em tempo real.



### Cálculo de Medidas

Calculamos médias, medianas, desvios-padrão e correlações de forma instantânea, mesmo processando milhões de registros simultaneamente.





# Etapa 5: Modelagem Preditiva e Inferência

É onde os conceitos estatísticos são implementados como algoritmos computacionais. A Informática é a "*cozinha*" onde a receita da Estatística é executada em grande escala, gerando modelos preditivos poderosos.



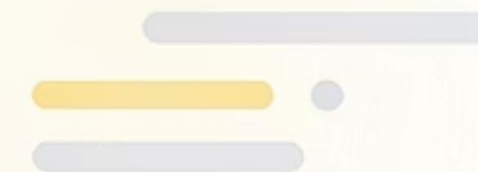
## Machine Learning

Algoritmos como Regressão Linear, Árvores de Decisão e Redes Neurais são modelos estatísticos implementados em código.



## Testes de Hipóteses

Sistemas de e-commerce usam testes A/B para decidir qual versão gera mais conversões, automatizando análises.



# Machine Learning: A Personificação da União

O **Aprendizado de Máquina** é a área que personifica perfeitamente essa união simbiótica. Algoritmos complexos são, em sua essência, modelos estatísticos sofisticados implementados em código para que computadores possam "*aprender*" padrões a partir dos dados.

## Regressão Linear

Previsão de valores contínuos baseada em relações lineares entre variáveis

## Regressão Logística

Classificação binária usando probabilidades para tomada de decisão

## Árvores de Decisão

Modelos hierárquicos que dividem dados em grupos homogêneos

## Redes Neurais

Arquiteturas complexas inspiradas no cérebro humano para padrões não-lineares

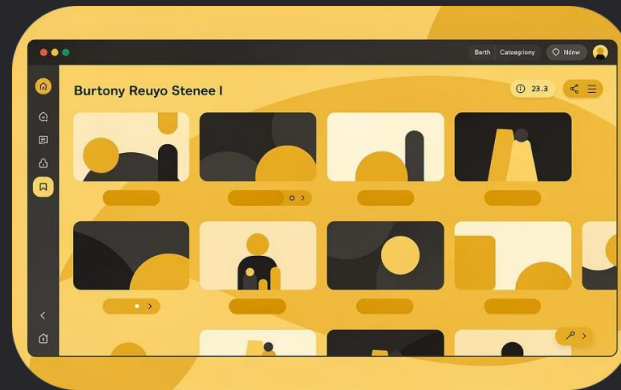
# Etapa 6: Tomada de Decisão

Os resultados das análises estatísticas são incorporados em sistemas computacionais para gerar valor real e impactar decisões de negócio em tempo real.



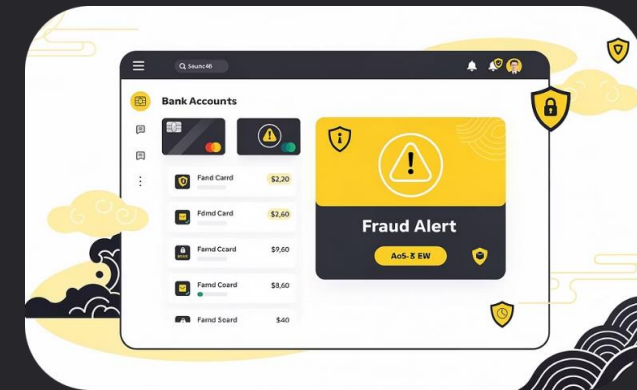
## Dashboards Interativos

Painéis criados com Power BI, Tableau ou bibliotecas web que mostram indicadores-chave (KPIs) em tempo real.



## Sistemas de Recomendação

A "magia" por trás das sugestões da Netflix, Amazon ou Spotify usando modelos estatísticos e computacionais.



## Detecção de Anomalias

Sistemas bancários que detectam transações suspeitas em milissegundos usando modelos estatísticos.

# O Combustível da Relação: Dados em Sistemas

Os dados são a matéria-prima fundamental que conecta a Estatística e a Informática. Em sistemas computacionais modernos, o uso de dados pode ser classificado em algumas categorias principais:



Cada tipo de dado requer abordagens específicas de armazenamento, processamento e análise estatística.



# Categorias de Dados em Sistemas

## Dados Operacionais (OLTP)

Dados do dia a dia gerados por cada transação: vendas, cadastros, cliques. Sistemas como PostgreSQL e MySQL são especialistas em processamento rápido e consistente.

## Big Data

Volumes, velocidade e variedade massivos que exigem tecnologias como Hadoop e Spark para processamento distribuído e aplicação de métodos estatísticos em escala.

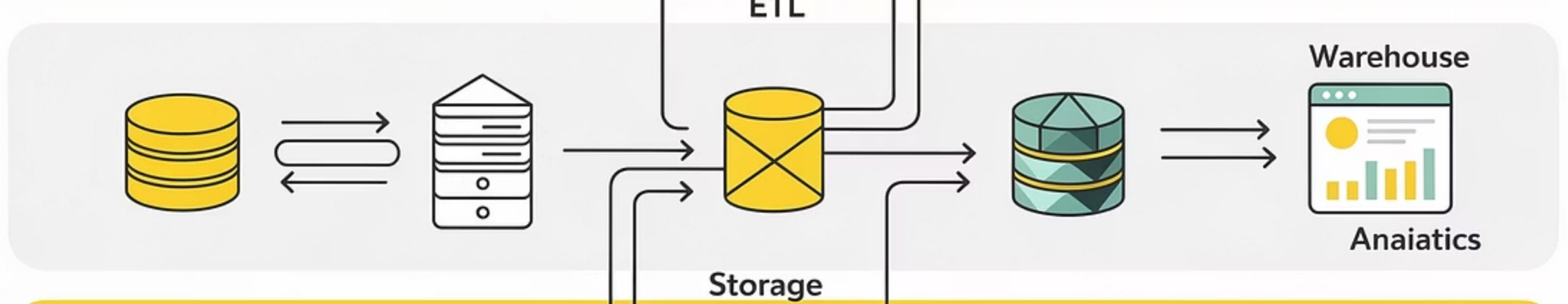
1

2

3

## Dados Analíticos (OLAP)

Dados históricos consolidados e preparados para análise. Armazenados em Data Warehouses (Amazon Redshift, Google BigQuery) para consultas complexas e agregações estatísticas.



# Data Warehouses: O Encontro Perfeito

Um Data Warehouse é, essencialmente, um grande banco de dados otimizado para consultas complexas e agregações — somas, médias, contagens — que são o **pão com manteiga da Estatística**.

## Características Principais:

- Dados históricos e consolidados de múltiplas fontes
- Otimizado para leitura e análises complexas
- Suporte a consultas OLAP e Business Intelligence
- Esquemas dimensionais (fatos e dimensões)
- Processamento de grandes volumes com rapidez

## Por Que Importa?

Permite que profissionais apliquem técnicas estatísticas sofisticadas em dados empresariais históricos para identificar tendências, padrões e insights estratégicos.



# Profissionais do Futuro

A Estatística e a Informática são como duas faces da mesma moeda na era digital. A **Informática** cria a infraestrutura e as ferramentas para lidar com os dados, enquanto a **Estatística** fornece o método científico para transformar esses dados em conhecimento, previsões e decisões inteligentes.

📌 **Conclusão:** Um profissional que domina ambas as áreas — capaz de programar sistemas computacionais E aplicar métodos estatísticos rigorosos — é um dos mais valiosos e procurados no mercado de trabalho atual. A simbiose entre essas duas ciências continuará a moldar o futuro da tecnologia e da tomada de decisões baseada em dados.