

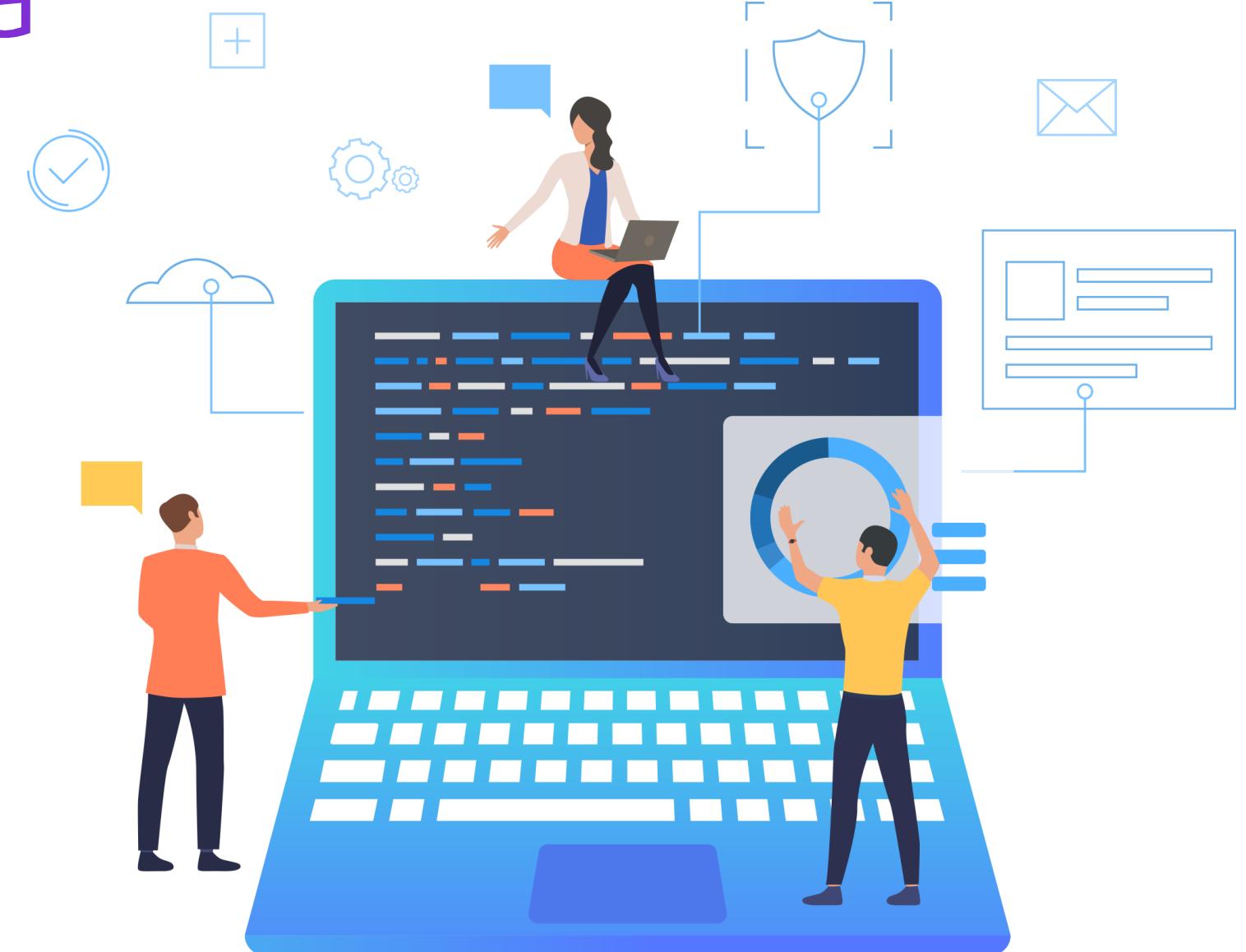


Análise de Dados e Big Data

Computação e Tecnologia da
Informação

01

Big Data e Analytics



Diego Augusto Barros

ENSINO & TECNOLOGIA

1. Big Data
2. Mineração de Dados
3. Tipos de Dados
4. Big Data Analytics
5. Ferramentas e Tecnologias





01

BIG DATA

INTRODUÇÃO

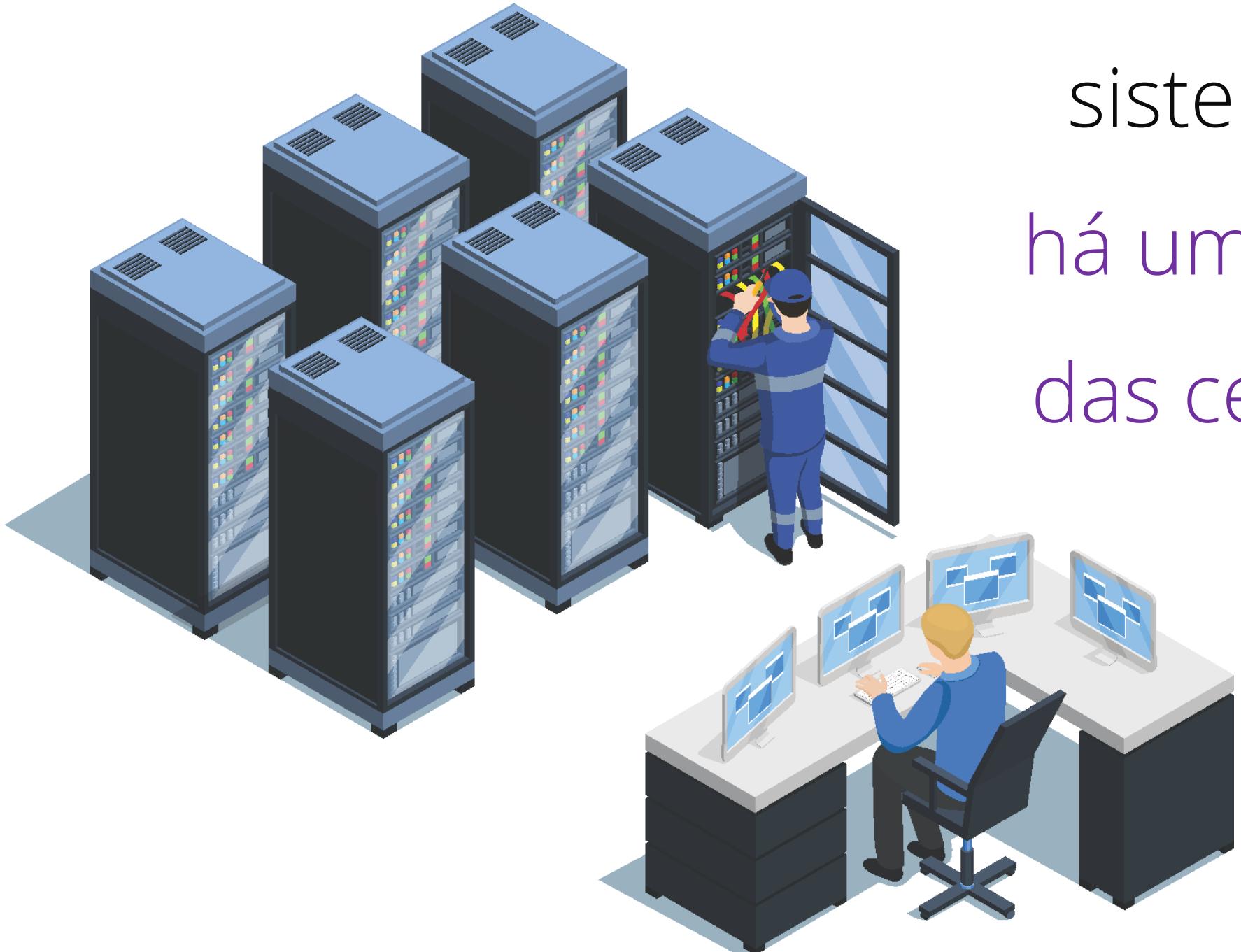


Os Bancos de Dados são essenciais na vida da sociedade moderna.

Diariamente, todos nós fazemos várias atividades que envolve alguma interação com Banco de Dados.

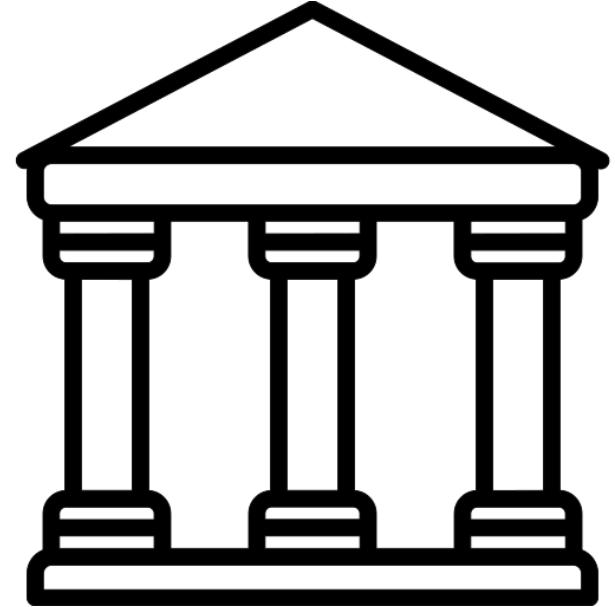


INTRODUÇÃO



Sempre que você usa algum sistema, muito provavelmente, há um Banco de Dados por trás das cenas, fornecendo todas as informações para você.

INTRODUÇÃO



Banco

Para depositar
ou fazer algum
saque.



Passagem Aérea

Reservar voo
em uma
companhia
aérea



Biblioteca

Para consultar
catálogo de
obras



Compras

Em sites de
comércio
eletrônico

INTRODUÇÃO



Todos os exemplos que vimos até agora são considerados **Sistemas Tradicionais de Banco de Dados**, pois, a maioria dos dados acessados estão no formato **numérico** ou **textual**.

INTRODUÇÃO

No entanto, com o avanço da tecnologia, o crescimento das Redes Sociais e sistemas cada vez mais interativos;



INTRODUÇÃO



As empresas estão colecionando cada vez mais quantidades de **dados não tradicionais**, diferentes de números e textos, tais como: publicações, tweets, imagens, jogos, vídeos e músicas.



INTRODUÇÃO

Dados não estruturados de outras fontes, como:

- Internet;
- Fotos;
- Vídeos;
- Áudio;
- Redes Sociais e sensores.

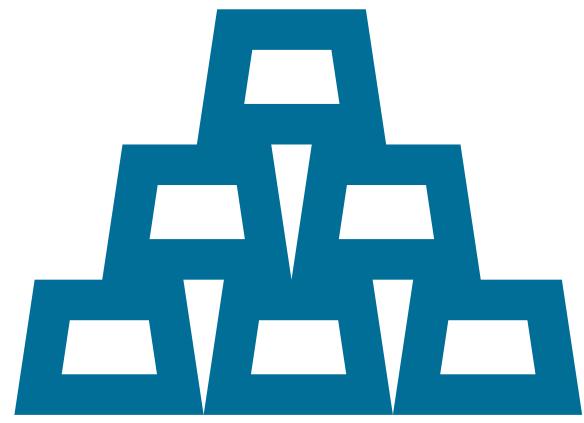


BIG DATA

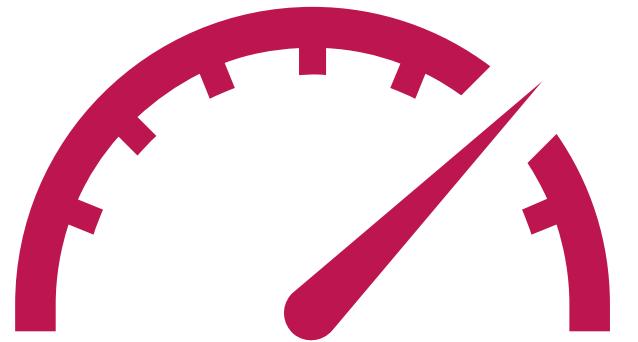


Grandes quantidades de dados não estruturados que são difíceis ou impossíveis de capturar, armazenar e manipular com os sistemas de gerenciamento de bancos de dados tradicionais.

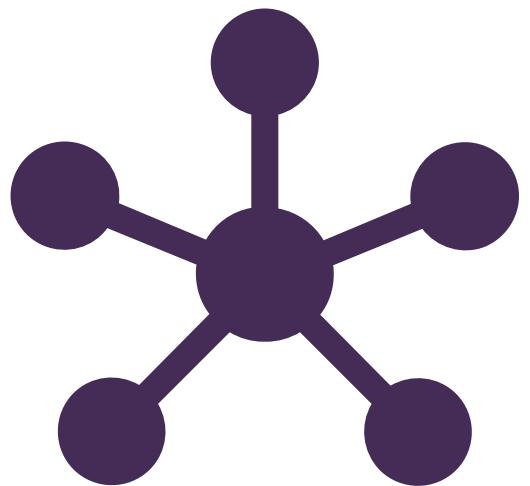
OS 3Vs do BIG DATA



VOLUME



VELOCIDADE

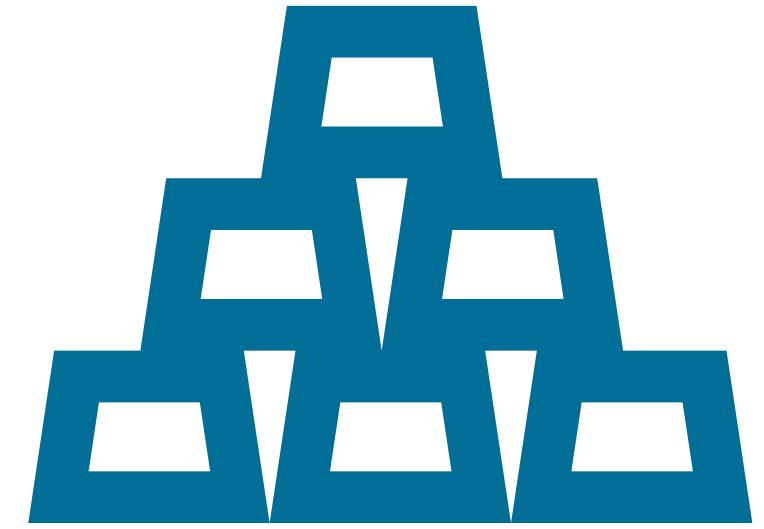


VARIEDADE



VOLUME

- Grandes volumes de dados não estruturados;
- Pode utilizar dezenas de terabytes de dados ou centenas de petabytes.

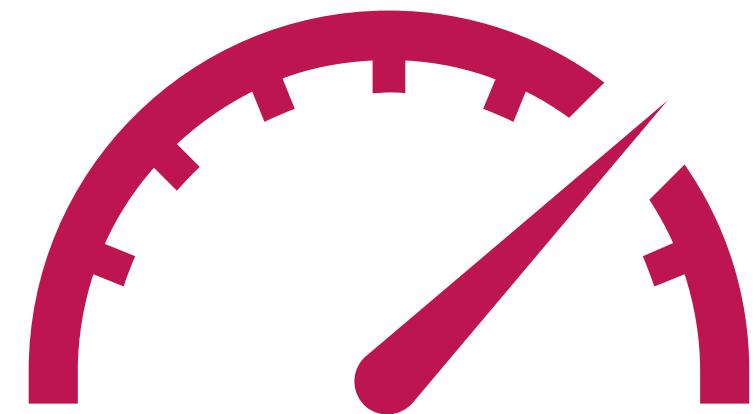


VOLUME



VELOCIDADE

- Desempenho do processamento de dados;
- Taxa na qual os dados são recebidos e talvez administrados.

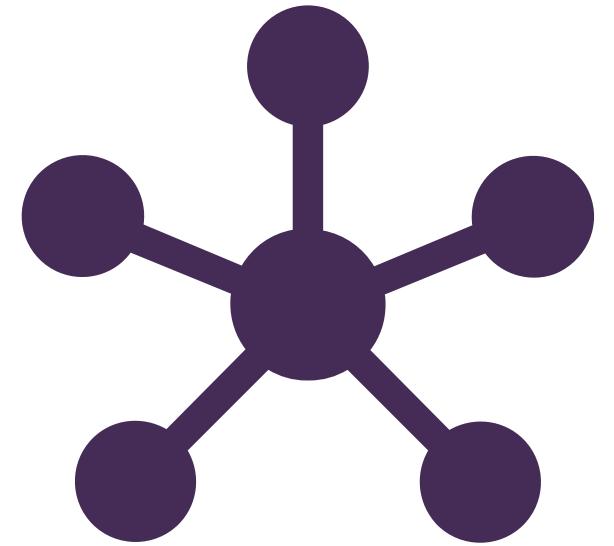


VELOCIDADE



VARIEDADE

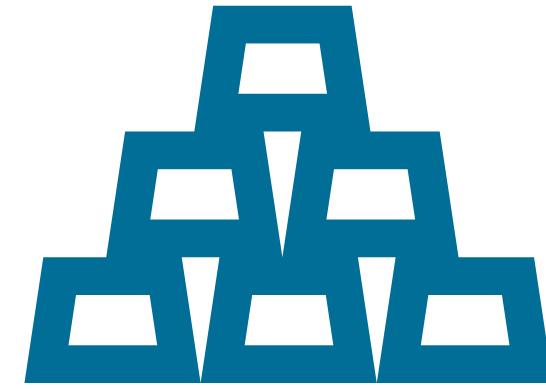
- Tipos de Dados Disponíveis
- Dados estruturados tradicionais
(números e textos)
- Dados não estruturados e
semiestruturados, como: texto, áudio
e vídeo



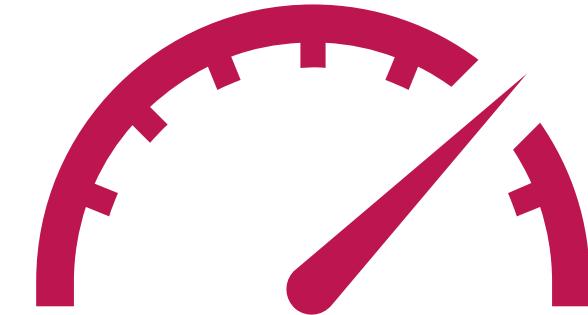
VARIEDADE



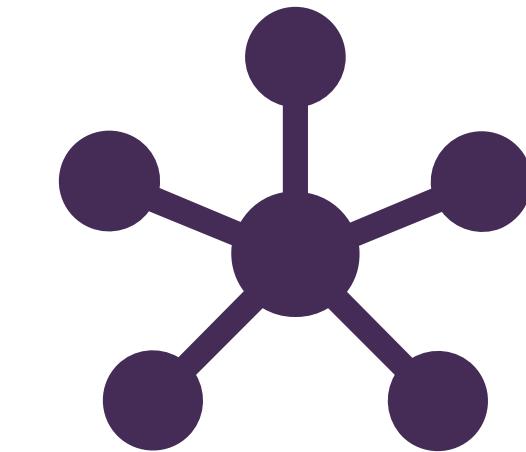
OS 5Vs do BIG DATA



VOLUME



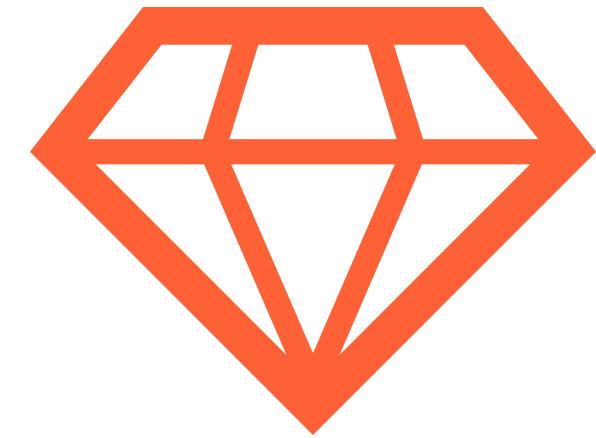
VELOCIDADE



VARIEDADE



VERACIDADE

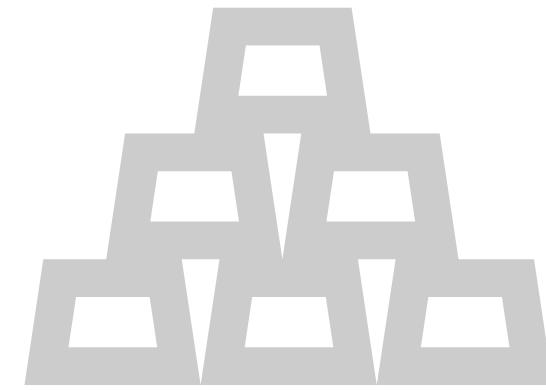


VALOR





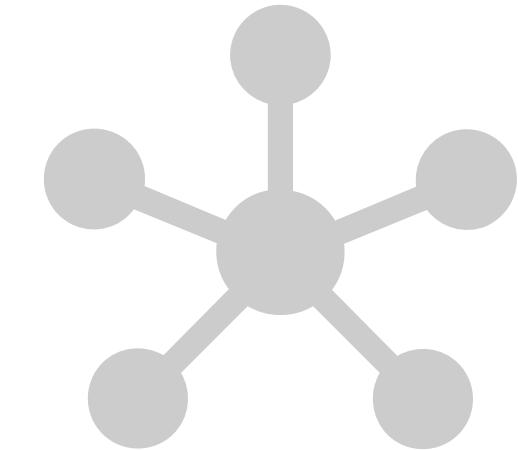
OS 5Vs do BIG DATA



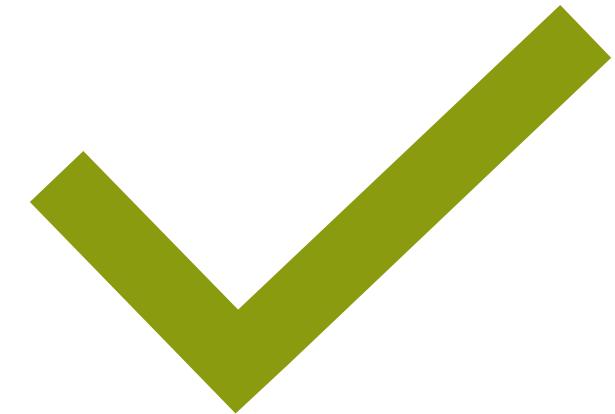
VOLUME



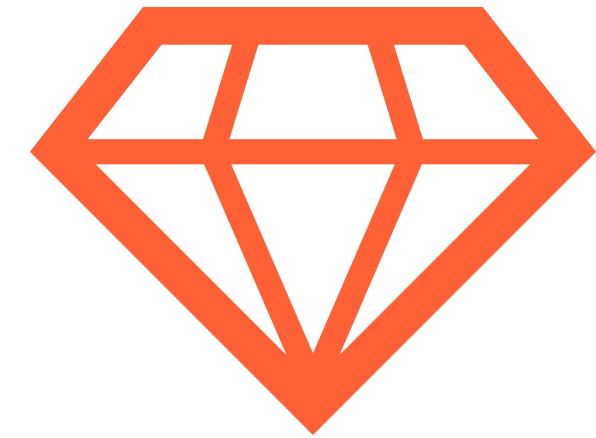
VELOCIDADE



VARIEDADE



VERACIDADE



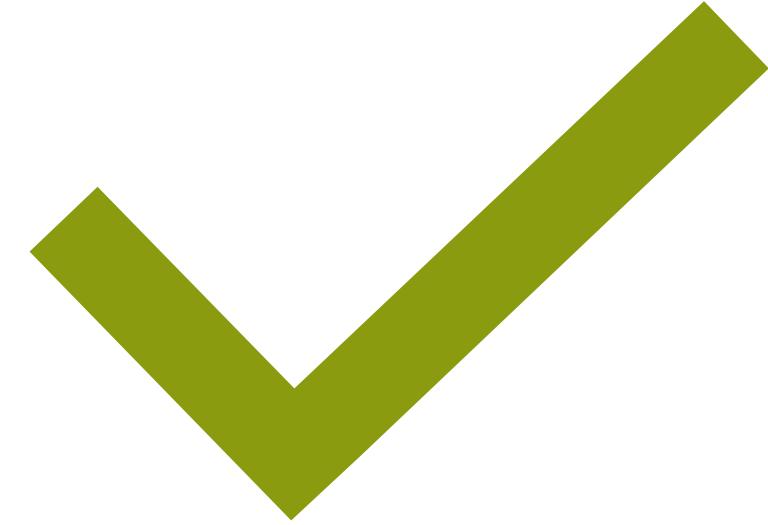
VALOR





VERACIDADE

O quanto verdadeira é a informação.



VERACIDADE

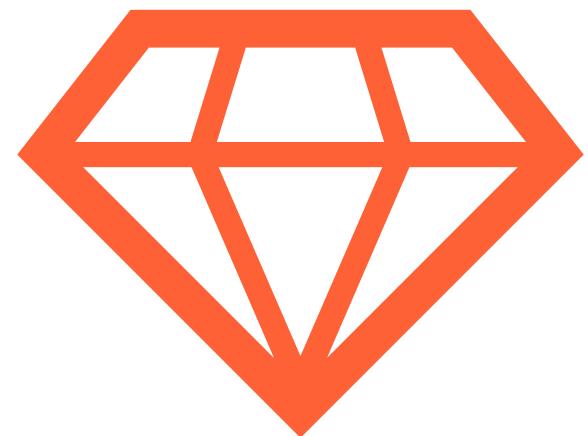
Quão confiáveis são seus dados, e quanto você pode confiar neles?



VALOR

O quanto útil é a informação

Grande parte do valor que as empresas oferecem vem de dados, que são analisados constantemente para produzir mais eficiência e desenvolver novos produtos.



VALOR

APLICAÇÕES PRÁTICAS

Manutenção Preventiva

- Prever falhas mecânicas podem estar profundamente relacionados a dados estruturados
 - Ano
 - Marca
 - Modelo



APLICAÇÕES PRÁTICAS



Manutenção Preventiva

- Análise de dados não estruturados
 - Milhões de entradas de log;
 - Dados de sensores;
 - Mensagens de erro;
 - Temperatura do motor.



APLICAÇÕES PRÁTICAS



Manutenção Preventiva

- Ao analisar os dados para encontrar possíveis problemas antes que eles ocorram;
- As empresas podem implementar a manutenção de maneira mais econômica;
- Maximizar vida útil de peças e equipamentos.



APLICAÇÕES PRÁTICAS



Fraude e Conformidade

- Identificar padrões em dados que indicam fraudes;
- Agregar grandes volumes de informações;
- Gerar os relatórios regulamentares mais eficientes e eficazes.



APLICAÇÕES PRÁTICAS



Eficiência Operacional

- Analisar e avaliar a produção, os comentários e as devoluções de clientes;
- Para **reduzir interrupções** e **anticipar demandas** futuras.





02

MINERAÇÃO DE DADOS

Data Mining

MINERAÇÃO DE DADOS



Mineração de Dados (*Data Mining*) é uma ferramenta de análise de informações que envolve a descoberta automatizada de padrões e relacionamentos em um conjunto de dados.

MINERAÇÃO DE DADOS



A mineração de dados é um processo de **extração e descoberta de padrões em grandes conjuntos de dados** envolvendo métodos que fazem interseção com:

- Aprendizado de Máquina;
- Estatística;
- Sistemas de Banco de Dados.



BLUE CROSS E BLUE SHIELD

“A criação de uma base de dados com 3,5 bilhões de informações foi a parte fácil. A **parte difícil** é **transformar** tudo isso em informações práticas e **insights** que melhorarão a qualidade e a realização dos serviços de cuidados, em última análise, dobrando a curva dos custos nos Estados Unidos” (CEO da Blue Cross e Blue Shield)



BlueCross[®]
BlueShield[®]

MINERAÇÃO DE DADOS

“Assim como em um garimpo, o *data mining* filtra montanhas de dados para encontrar algumas pepitas de informações valiosas.”

(LAUDON, 2011)



MINERAÇÃO DE DADOS

O objetivo é a extração de padrões e conhecimento de grandes quantidades de dados, não a extração (mineração) de dados em si.



ANÁLISE DE DADOS X DATA MINING



Análise de Dados

Usada para testar modelos e hipóteses no conjunto de dados, por exemplo, analisando a eficácia de uma campanha de marketing, independentemente da quantidade de dados

Mineração de Dados (Data Mining)

A mineração de dados usa aprendizado de máquina e modelos estatísticos para descobrir padrões ocultos em um grande volume de dados



EXEMPLO: BROOKS BROTHERS

O mais antigo varejo de roupas dos EUA, utilizou o **data mining** para fornecer relatórios aos gestores da loja para melhorar o desempenho da loja e a **satisfação** dos clientes





EXEMPLO: INSTITUIÇÕES MÉDICAS

Instituições de saúde utilizam o **data mining** para ajudar médicos a analisarem os grandes bancos de dados e assim melhorarem os diagnósticos e os tratamentos dos pacientes.



ANÁLISE PREDITIVA

Forma de **data mining** que combina os **dados históricos** com suposições sobre futuras condições para prever resultados de eventos, como **futuras vendas de produtos** ou a probabilidade de que um cliente ficará **inadimplente** em um empréstimo.

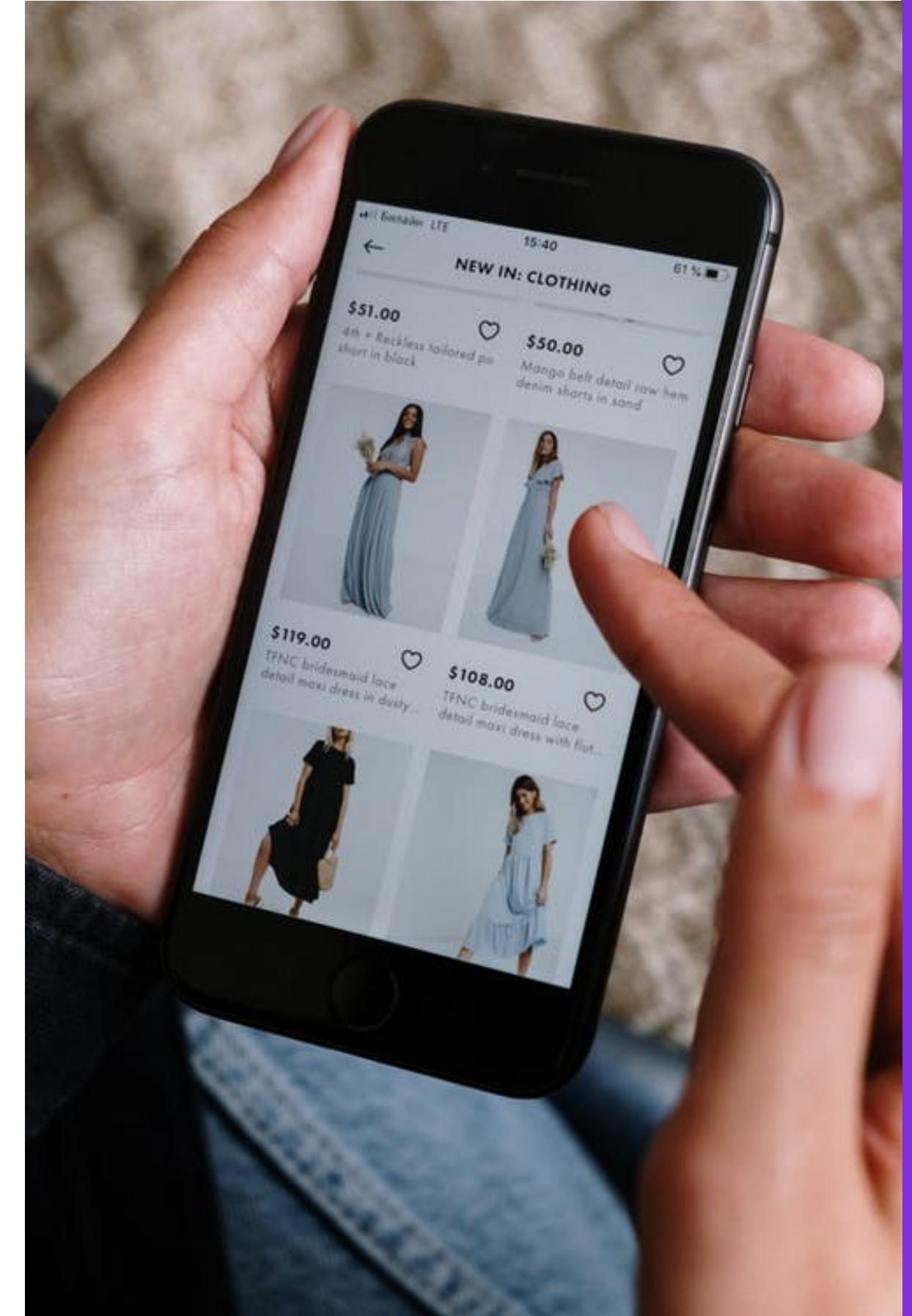




EXEMPLO: MERCADO VAREJISTA

Os varejistas utilizam a análise preditiva para:

- Promover clientes ocasionais em compradores frequentes;
- Prever quais produtos eles comprarão se lhes forem oferecidos incentivos apropriados



ANÁLISE PREDITIVA

Também chamada de Analytics por alguns, a análise preditiva sempre utiliza técnicas matemáticas sofisticadas para analisar vastos depósitos de dados.



ANÁLISE PREDITIVA

“Só ter uma pilha de dados não é tão útil. Você tem de fazer as perguntas certas e processar os dados da maneira correta onde puder fazer análises significativas”

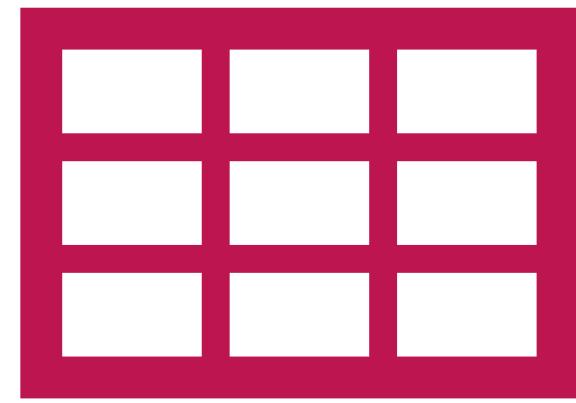




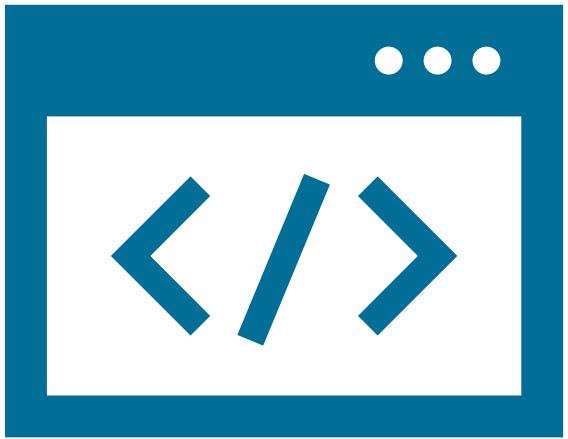
03

TIPOS DE DADOS

TIPOS DE DADOS



ESTRUTURADOS



SEMIESTRUTURADOS



SEMIESTRUTURADOS



DADOS ESTRUTURADOS

Dados que obedecem à estrutura tabular de modelos de dados, associados a bancos de dados relacionais ou outras formas de tabelas de dados.

	<code>id_vendedor</code>	<code>id_cidade</code>	<code>nome_vendedor</code>	<code>comissao</code>	<code>id_cliente</code>	<code>data venda</code>	<code>valor_yenda</code>
▶	1	6	Bernardo	0.10	2	2020-06-16	700.00
	2	1	Giovana	0.25	1	2019-08-15	501.10
	2	1	Giovana	0.25	4	2019-08-16	645.30
	2	1	Giovana	0.25	1	2019-09-02	580.00
	2	1	Giovana	0.25	4	2019-09-06	600.25
	2	1	Giovana	0.25	1	2019-09-15	406.24
	2	1	Giovana	0.25	1	2019-09-17	560.00
	3	7	Nilton	0.15	7	2019-10-09	560.30
	3	7	Nilton	0.15	7	2019-10-28	250.45
	4	1	Paulo	0.25	1	2019-10-16	576.00
	4	1	Paulo	0.25	1	2019-10-18	343.86
	4	1	Paulo	0.25	8	2019-10-27	675.00
	5	2	Vanessa	0.20	9	2019-08-15	600.00
	5	2	Vanessa	0.20	9	2019-08-22	520.25
	5	2	Vanessa	0.20	5	2019-10-14	505.10
	5	2	Vanessa	0.20	5	2019-10-19	480.25

Dados formatados e organizados em tabelas (linhas e colunas)



DADOS SEMIESTRUTURADOS

Forma de dados estruturados que
não obedece à estrutura tabular de
modelos de dados associados a
bancos de dados relacionais ou
outras formas de tabelas de dados.

```
1 {  
2   _id: "19f86359f6c111ea9c830c9d",  
3   nome: "Douglas",  
4   sobrenome: "Crockford",  
5   contato: {  
6     email: "douglas@crockford.com",  
7     website: "https://www.crockford.com/"  
8   }  
9 }
```

```
1 <catalogo>  
2   <livro id="livro-1">  
3     <autor>Gambardella, Matthew</autor>  
4     <titulo>XML Developer's Guide</titulo>  
5     <genero>Computer</genero>  
6     <preco>44.95</preco>  
7     <data_publicacao>2000-10-01</data_publicacao>  
8     <descricao>An in-depth look at creating applications  
9       with XML.</descricao>  
10    </livro>  
11 </catalogo>
```



DADOS SEMIESTRUTURADOS

Contém **tags** ou outros marcadores para separar elementos semânticos e impor hierarquias de registros e campos nos dados.

```
1 {  
2   _id: "19f86359f6c111ea9c830c9d",  
3   nome: "Douglas",  
4   sobrenome: "Crockford",  
5   contato: {  
6     email: "douglas@crockford.com",  
7     website: "https://www.crockford.com/"  
8   }  
9 }
```

```
1 <catalogo>  
2   <livro id="livro-1">  
3     <autor>Gambardella, Matthew</autor>  
4     <titulo>XML Developer's Guide</titulo>  
5     <genero>Computer</genero>  
6     <preco>44.95</preco>  
7     <data_publicacao>2000-10-01</data_publicacao>  
8     <descricao>An in-depth look at creating applications  
9       with XML.</descricao>  
10    </livro>  
11 </catalogo>
```



EXEMPLO: DADOS SEMIESTRUTURADOS



```
1 <catalogo>
2   <livro id="livro-1">
3     <autor>Gambardella, Matthew</autor>
4     <titulo>XML Developer's Guide</titulo>
5     <genero>Computer</genero>
6     <preco>44.95</preco>
7     <data_publicacao>2000-10-01</data_publicacao>
8     <descricao>An in-depth look at creating applications
9       with XML.</descricao>
10    </livro>
11 </catalogo>
```



```
1 {
2   _id: "19f86359f6c111ea9c830c9d",
3   nome: "Douglas",
4   sobrenome: "Crockford",
5   contato: {
6     email: "douglas@crockford.com",
7     website: "https://www.crockford.com/"
8   }
9 }
```

ARQUIVO XML

ARQUIVO JSON

DADOS SEMIESTRUTURADOS: XML



```
1 <catalogo>
2   <livro id="livro-1">
3     <autor>Gambardella, Matthew</autor>
4     <titulo>XML Developer's Guide</titulo>
5     <genero>Computer</genero>
6     <preco>44.95</preco>
7     <data_publicacao>2000-10-01</data_publicacao>
8     <descricao>An in-depth look at creating applications
9       with XML.</descricao>
10    </livro>
11 </catalogo>
```



DADOS SEMIESTRUTURADOS: JSON

```
1 {  
2   _id: "19f86359f6c111ea9c830c9d",  
3   nome: "Douglas",  
4   sobrenome: "Crockford",  
5   contato: {  
6     email: "douglas@crockford.com",  
7     website: "https://www.crockford.com/"  
8   }  
9 }
```



DADOS NÃO ESTRUTURADOS

Informações que não possuem um modelo de dados pré-definido ou não estão organizadas de forma pré-definida.



- Jogos;
- Vídeos;
- Imagens;
- Áudio e Músicas;
- Dados de Sensores;
- Mensagens de e-mails;
- Documentos de Texto;
- Mensagens de Redes Sociais.





04

BIG DATA ANALYTICS

Análise de Big Data



BIG DATA ANALYTICS



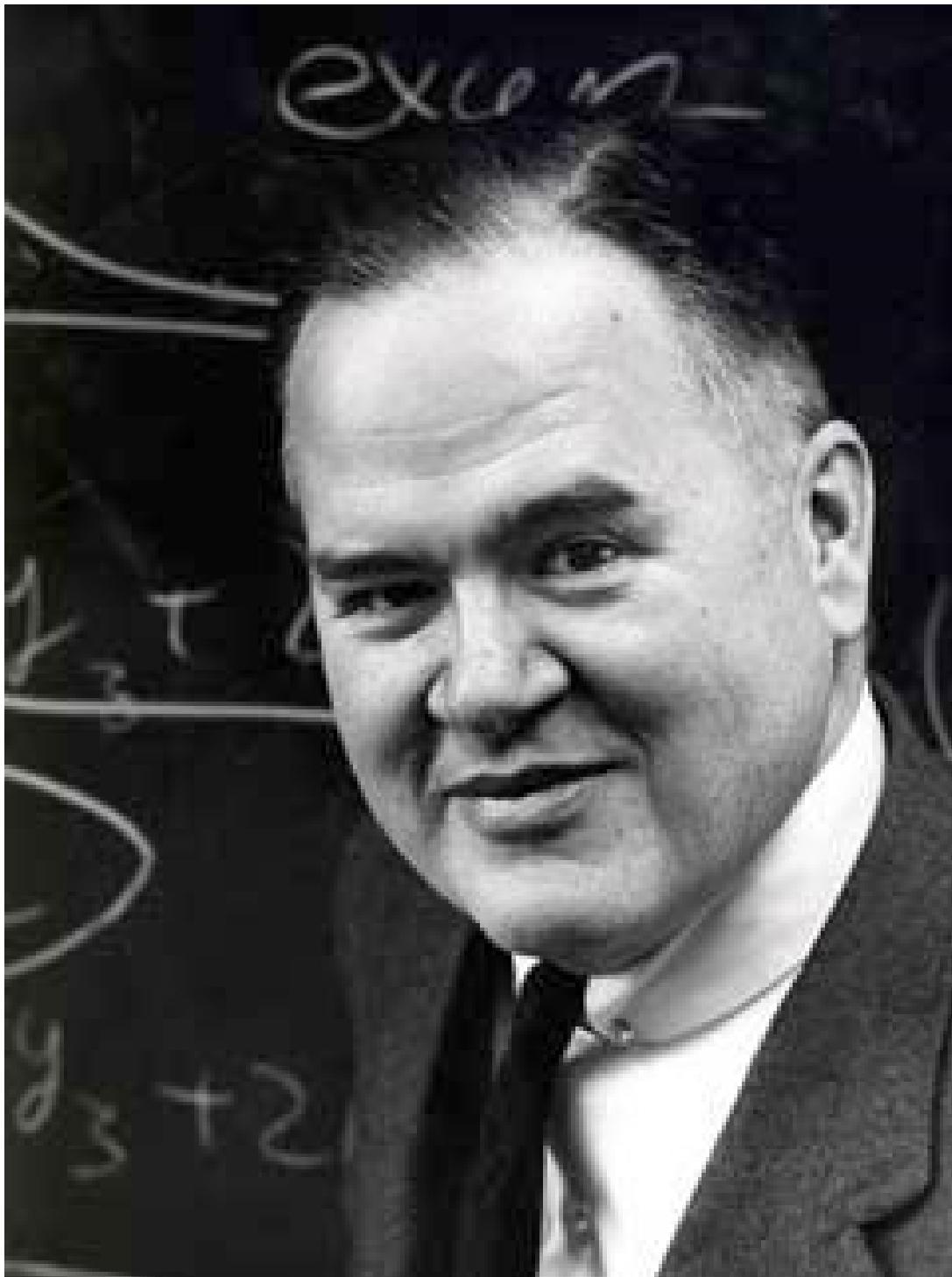
A análise de Big Data (Big Data Analytics) é o **uso de técnicas analíticas avançadas em conjuntos de big data muito grandes e diversos** que incluem dados estruturados, semiestruturados e não estruturados, de diferentes fontes e em diferentes tamanhos de terabytes a zetabytes.

DEFINIÇÃO DE ANÁLISE DE DADOS

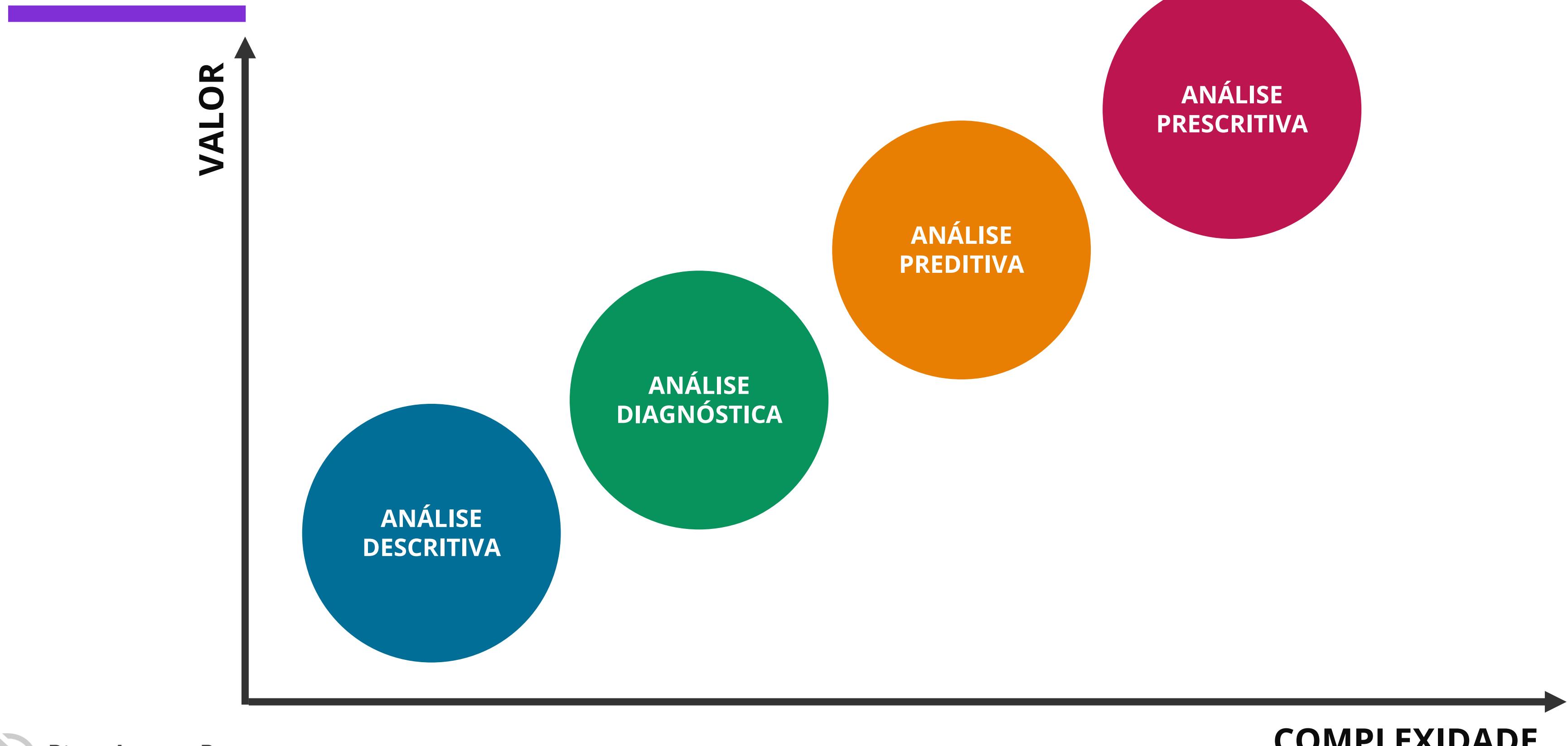


Procedimentos de análise de dados, **técnicas de interpretação dos resultados** de tais procedimentos, **formas de planejar a coleta de dados** para tornar sua análise mais fácil, mais precisa ou mais exata, e todo o maquinário e **resultados das estatísticas** (matemáticas) aplicáveis à análise de dados.

(John Turkey, 1961)



TIPOS DE ANÁLISES DE DADOS





ANÁLISE DESCRIPTIVA

Sumário estatístico que descreve quantitativamente ou resume características de uma coleção de informações.

É etapa inicial da análise usada para **descrever e resumir os dados**

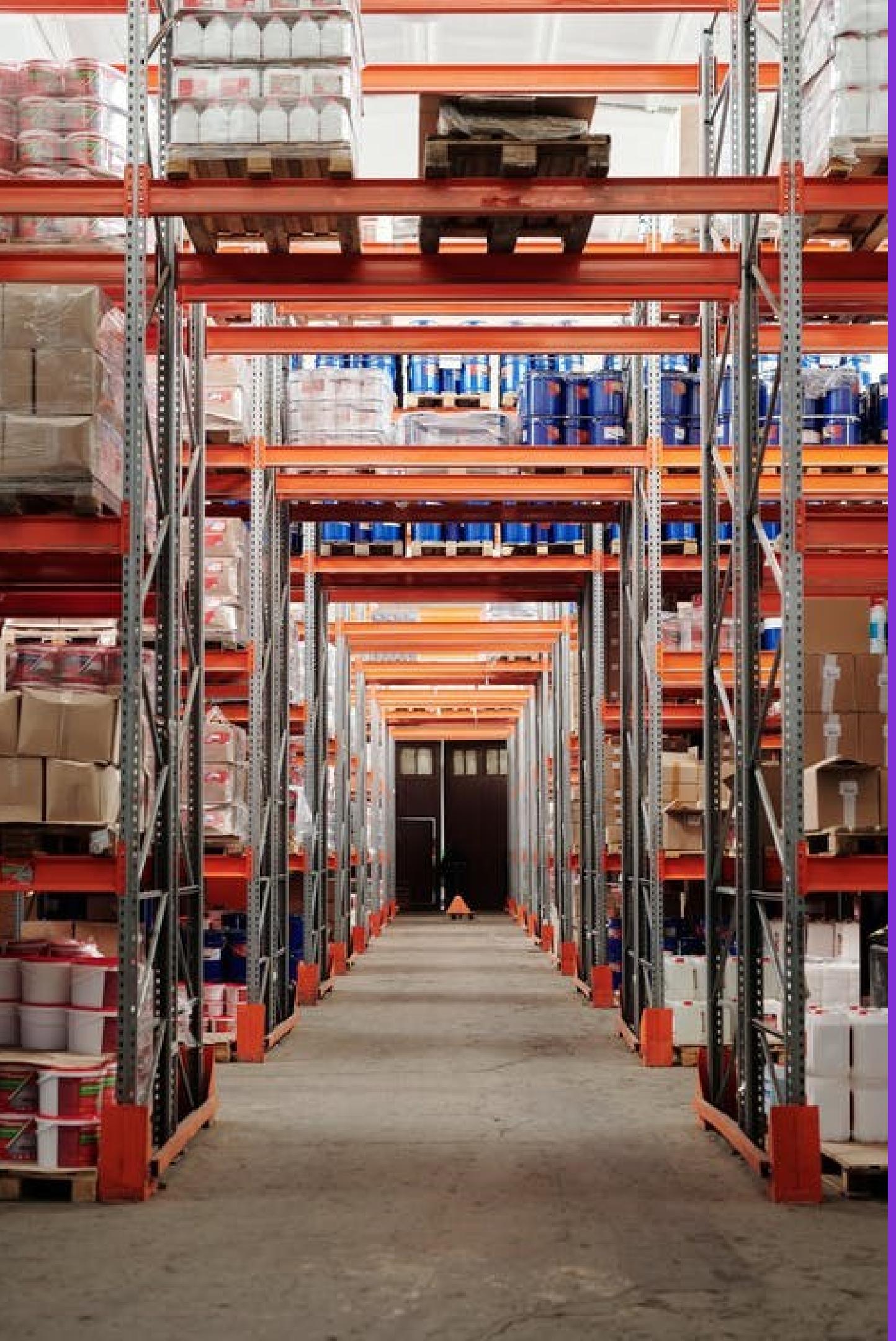


ANÁLISE
DESCRIPTIVA

O QUE ACONTECEU?

ANÁLISE DESCRIPTIVA

- Ao analisar a **receita mensal**, a **renda por grupo de produtos** e a **quantidade total de peças** de metal produzidas por mês.
- Um fabricante foi capaz de responder a uma série de perguntas sobre "o que aconteceu" e **decidir quais categorias de produtos** focar



ANÁLISE DIAGNÓSTICA

É a identificação da natureza e causa de um determinado fenômeno. Os **dados históricos** podem ser **analisados** em comparação com outros dados para responder à pergunta do por que algo aconteceu.



POR QUE ACONTECEU?



ANÁLISE DIAGNÓSTICA



- Através de um **dashboard** um varejista pode:
 - detalhar as vendas
 - lucro bruto em categorias
 - Para descobrir por que eles não alcançaram sua meta de lucro líquido.



ANÁLISE PREDITIVA

Combina os dados históricos com suposições sobre futuras condições para prever resultados de eventos.

Pode usar as descobertas das análises descritivas e diagnósticas, para detectar clusters e exceções a fim de prever tendências futuras.



O QUE É PROVÁVEL
QUE ACONTEÇA?



ANÁLISE PREDITIVA



A Procter & Gamble desenvolveu um programa chamado **Business Sufficiency** (suficiência do negócio) para ajudá-la a prever até um ano de vendas futuras

P&G
Procter & Gamble



ANÁLISE PRESCRITIVA

- Objetivo: prescrever qual ação tomar para eliminar um problema futuro ou aproveitar ao máximo uma tendência promissora.
- Ferramentas: Aprendizado de Máquina, Regras de Negócio e Algoritmos Especializados.
- Alta complexidade para implementar e gerenciar devido as ferramentas

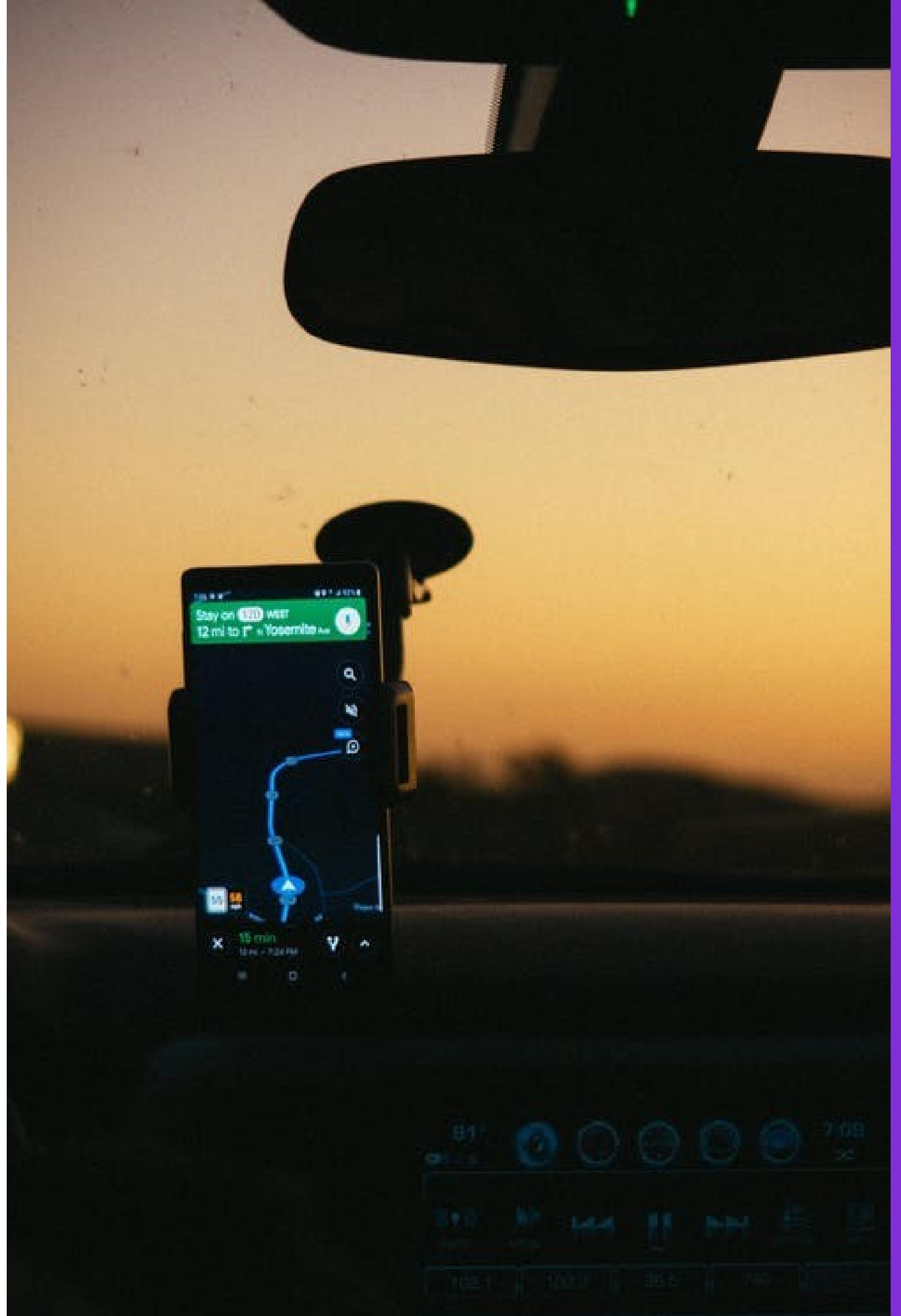


O QUE EU
PRECISO FAZER?



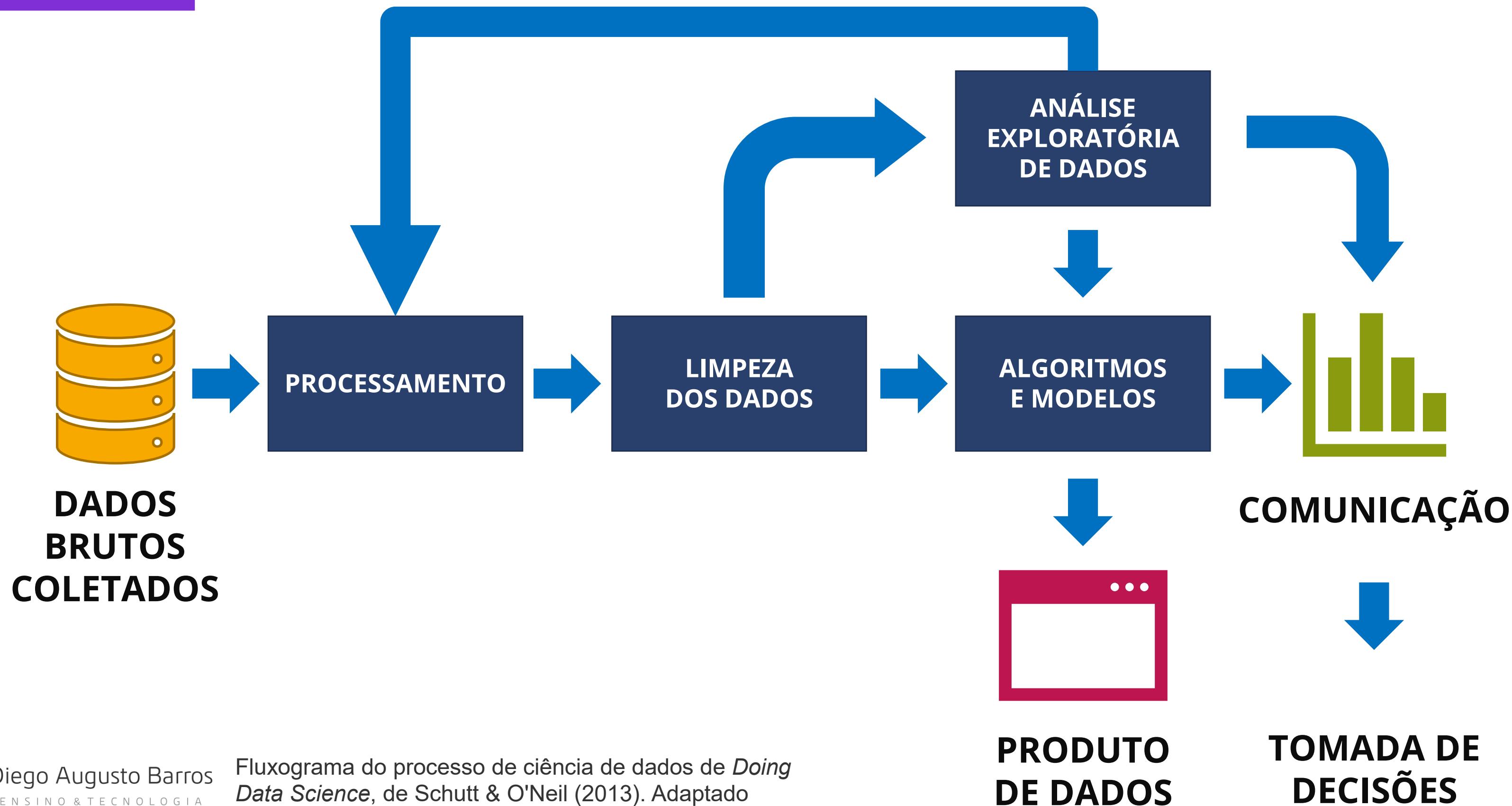
ANÁLISE PRESCRITIVA

- Um exemplo de análise prescritiva no nosso dia a dia é o **GPS**;
- Através de diferentes tipos de informações do trânsito e algoritmos ele prescreve a melhor rota para você eliminar o problema futuro de ficar preso no trânsito.





PROCESSO DE ANÁLISE DE DADOS

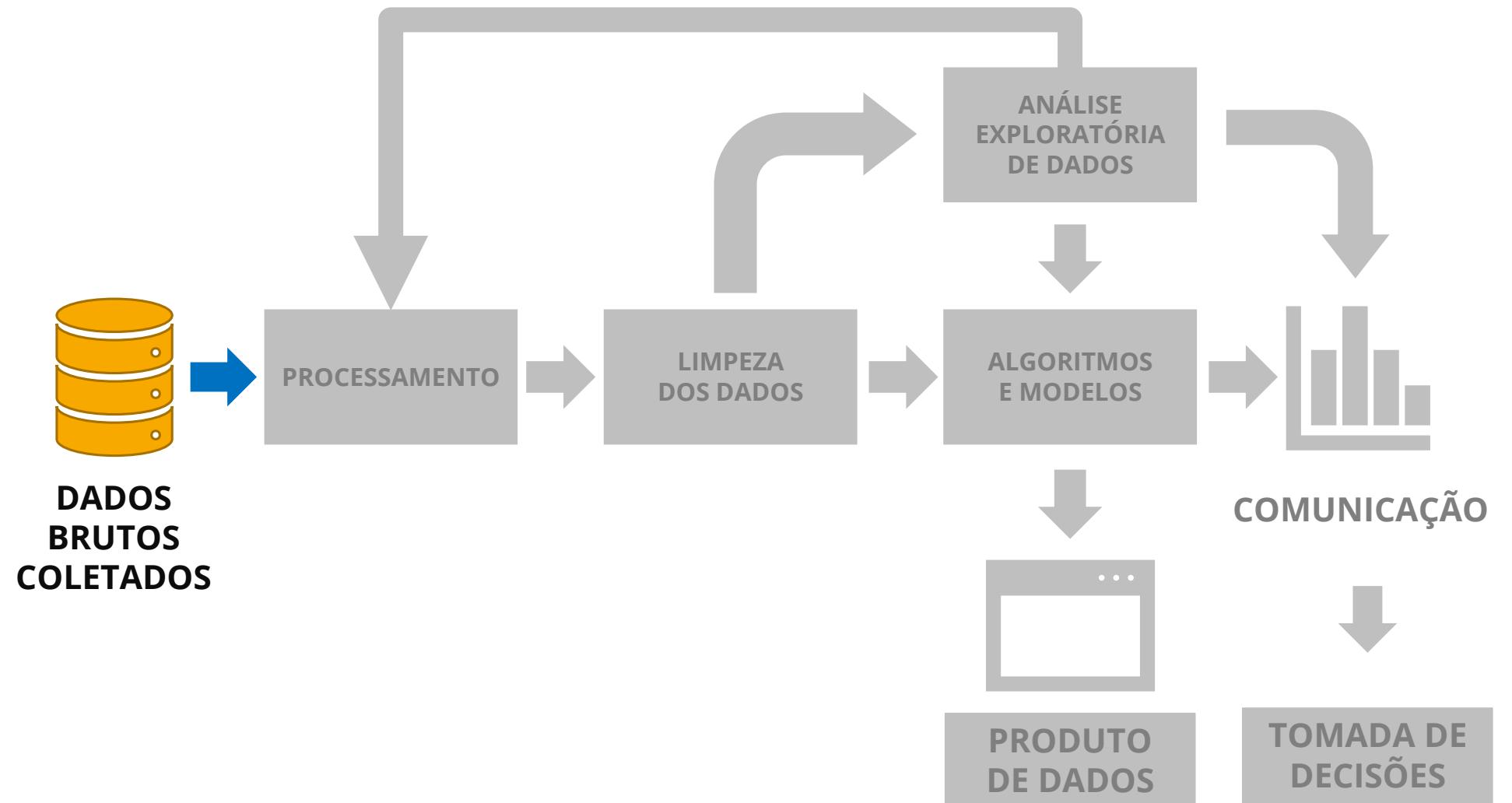




DADOS BRUTOS COLETADOS

Os dados são coletados de várias fontes

- Páginas da Web
- Sensores
- Câmeras
- Satélites
- Dispositivos de Gravação
- Textos
- Áudios

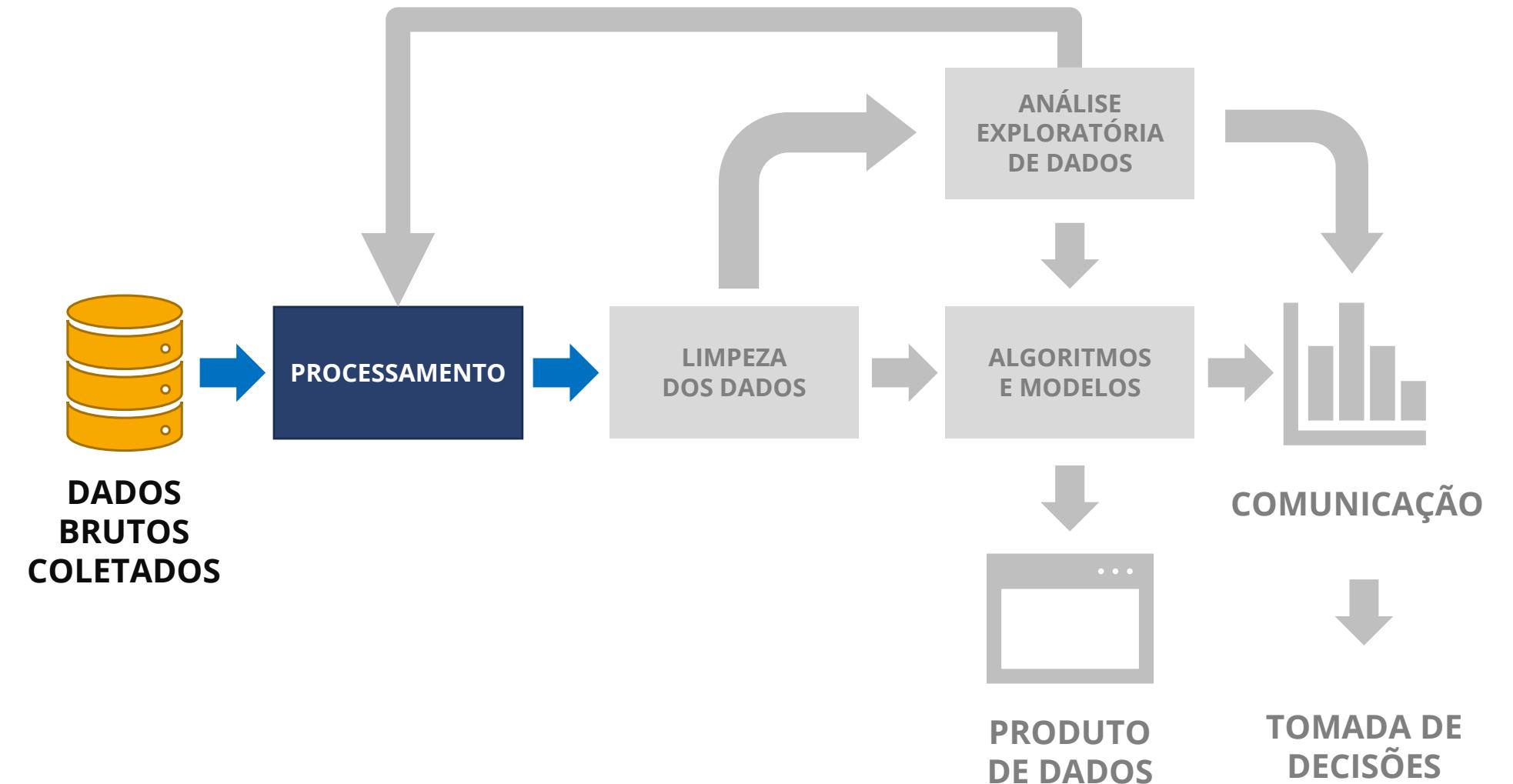




PROCESSAMENTO

Os dados são processados ou organizados para análise

- Por exemplo: formatados de maneira tabular, hierárquica ou em Redes

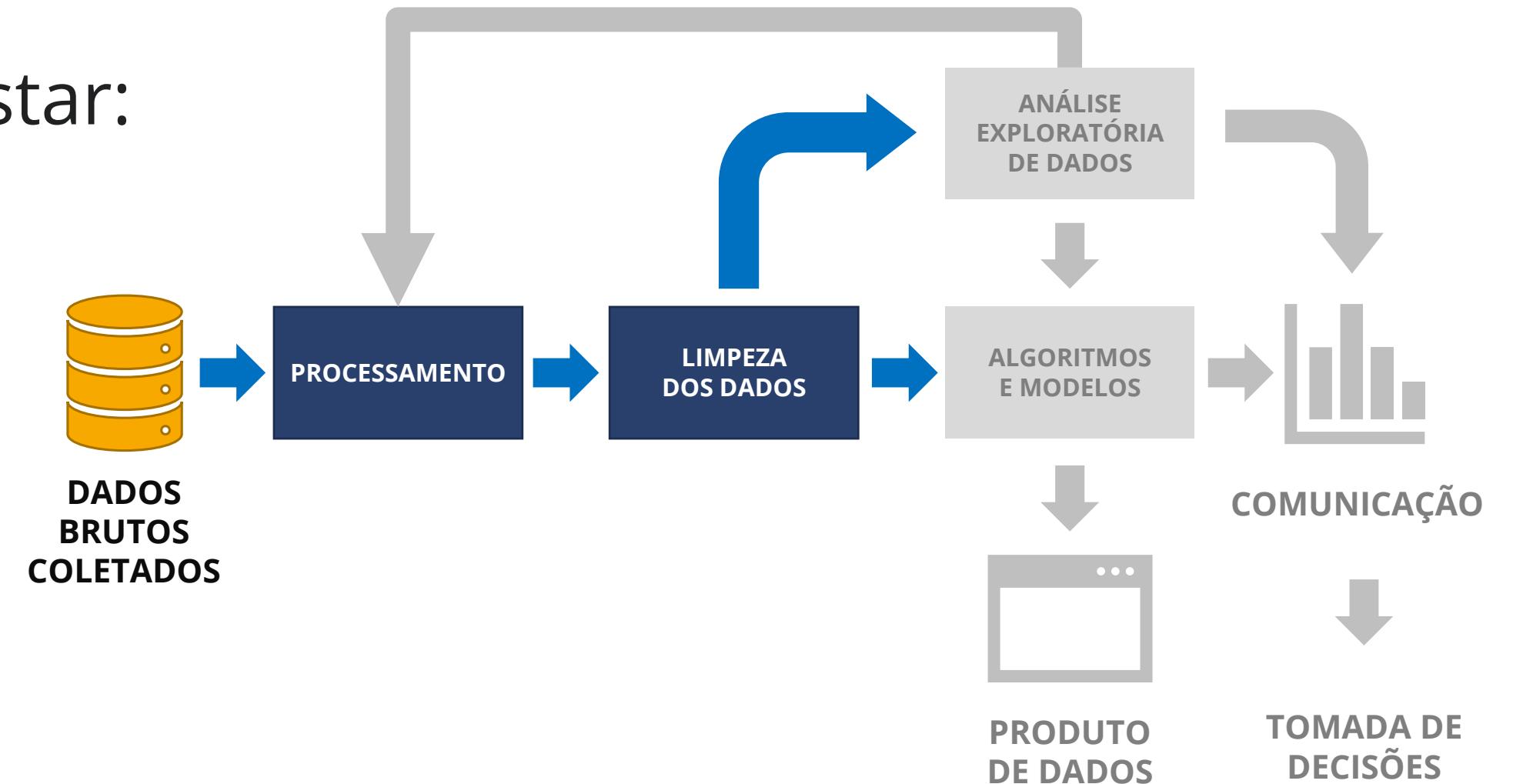




LIMPEZA DOS DADOS

Uma vez processados e organizados, os dados podem estar:

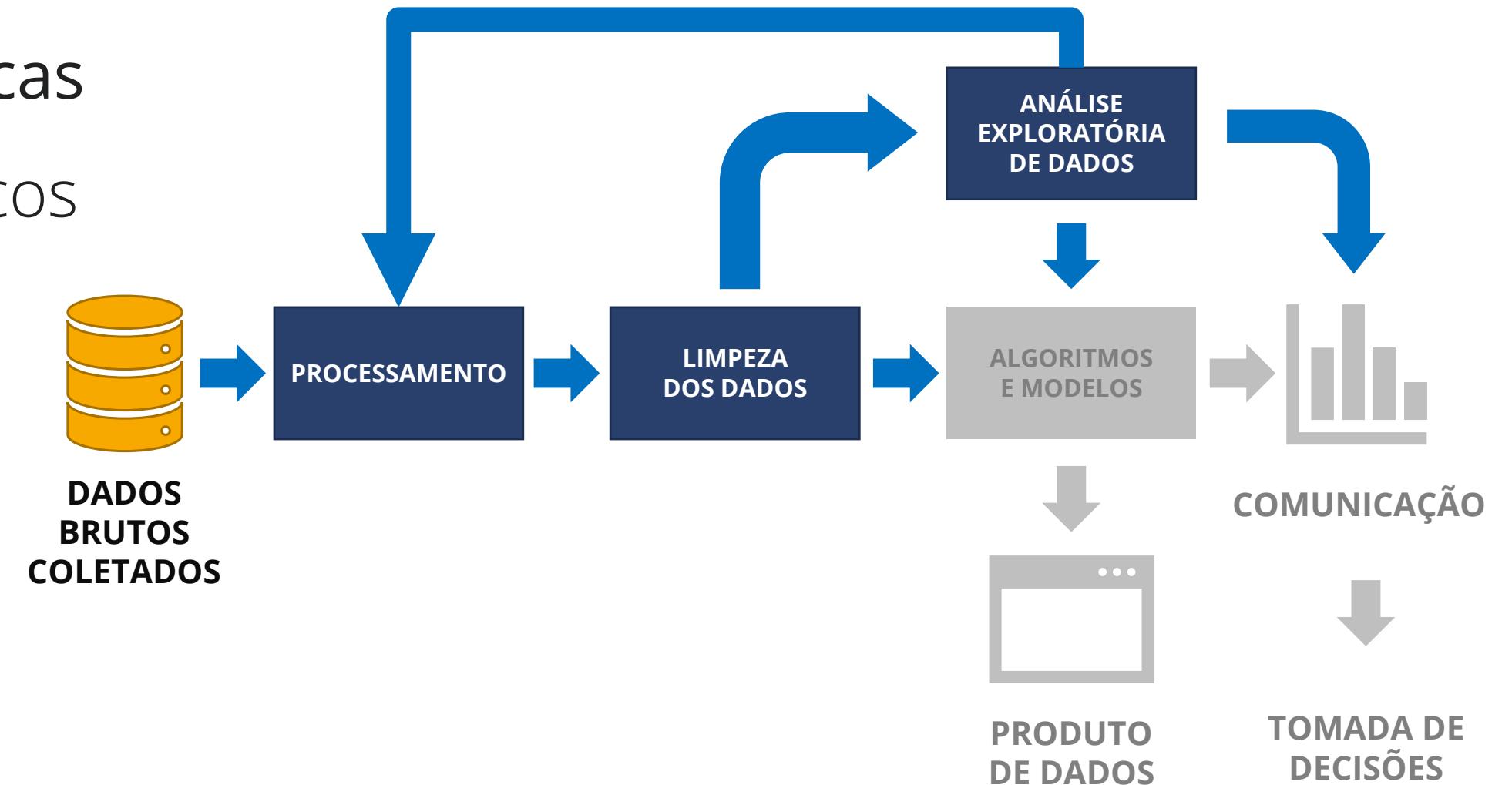
- Incompletos;
- Duplicados;
- Erros.





ANÁLISE EXPLORATÓRIA DOS DADOS

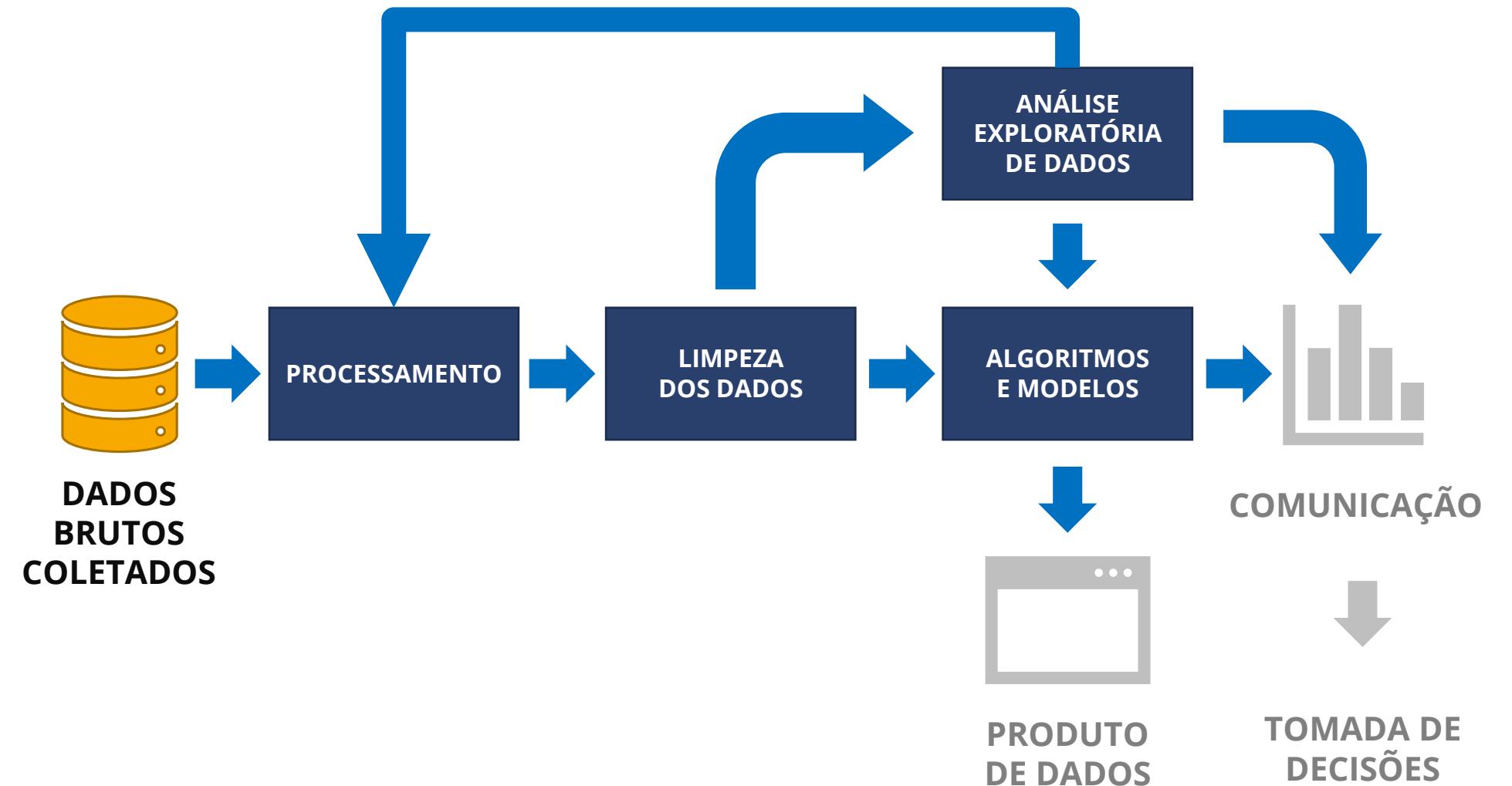
Abordagem de análise de conjuntos de dados para **resumir suas características principais**, muitas vezes usando gráficos estatísticos e outros métodos de visualização de dados





ALGORITMOS E MODELOS

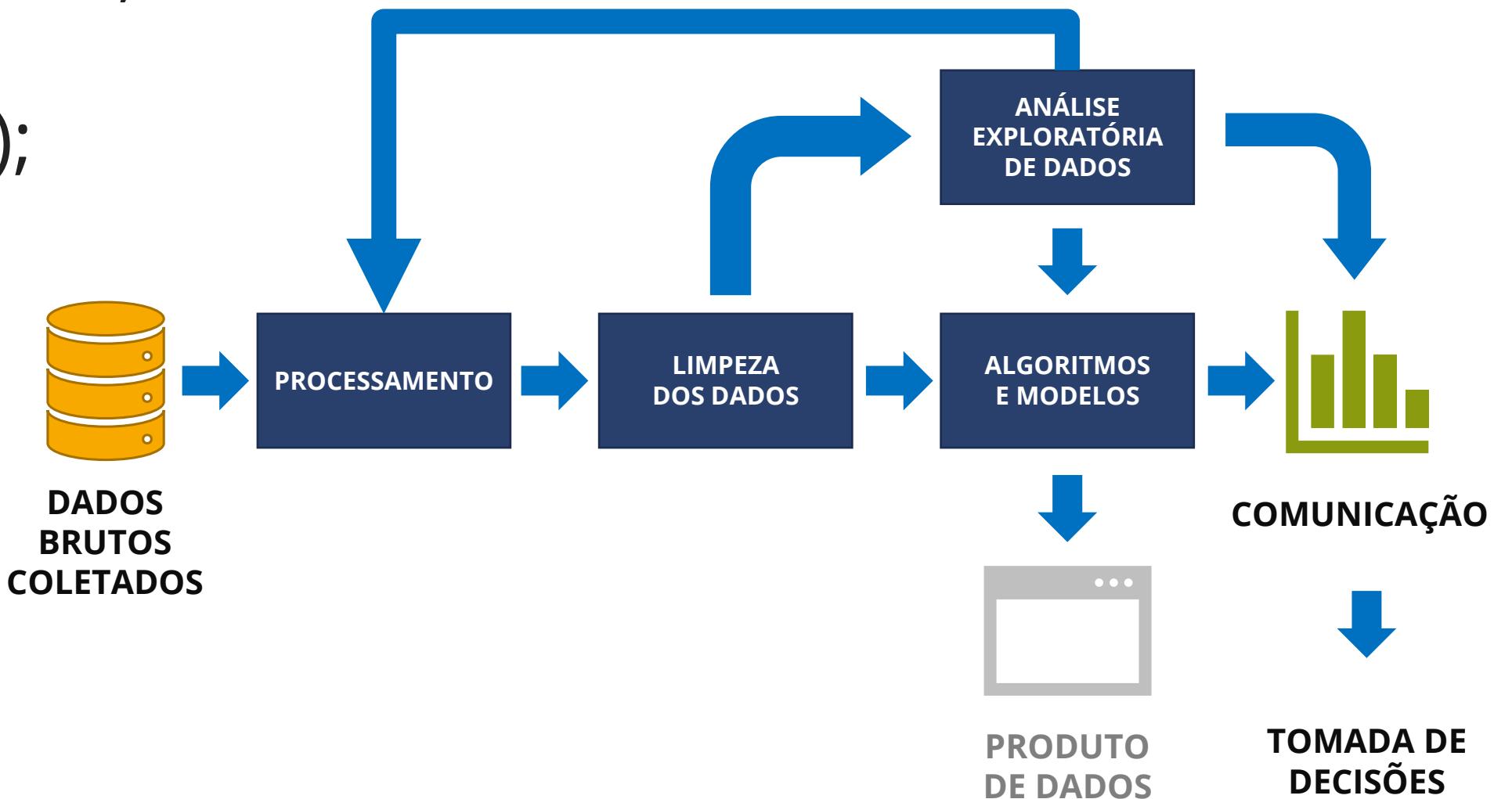
Técnicas de Inteligência Artificial que através do uso de modelos e algoritmos, são capazes de explorar um conjunto de dados, evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento.



COMUNICAÇÃO



- Técnicas de Visualização de Dados;
- Painéis Analíticos (Dashboards);
- Relatórios.

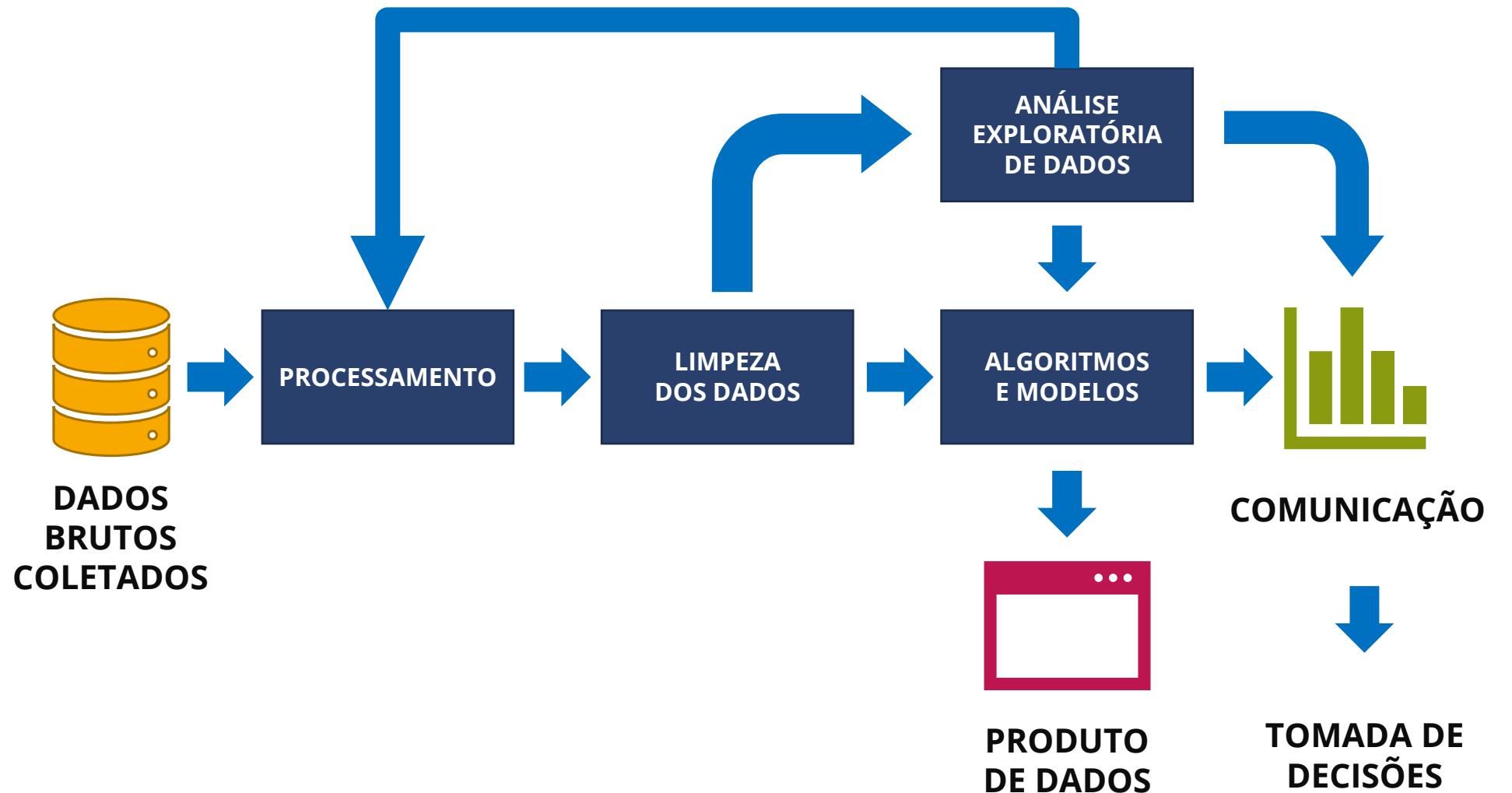




PRODUTO DE DADOS

Programa de computador que recebe entradas de dados e gera saídas, devolvendo-os ao ambiente.

- **Exemplo:** Software que analisa dados sobre o histórico de compras de um cliente e usa os resultados para recomendar novos produtos





05

FERRAMENTAS E TECNOLOGIAS

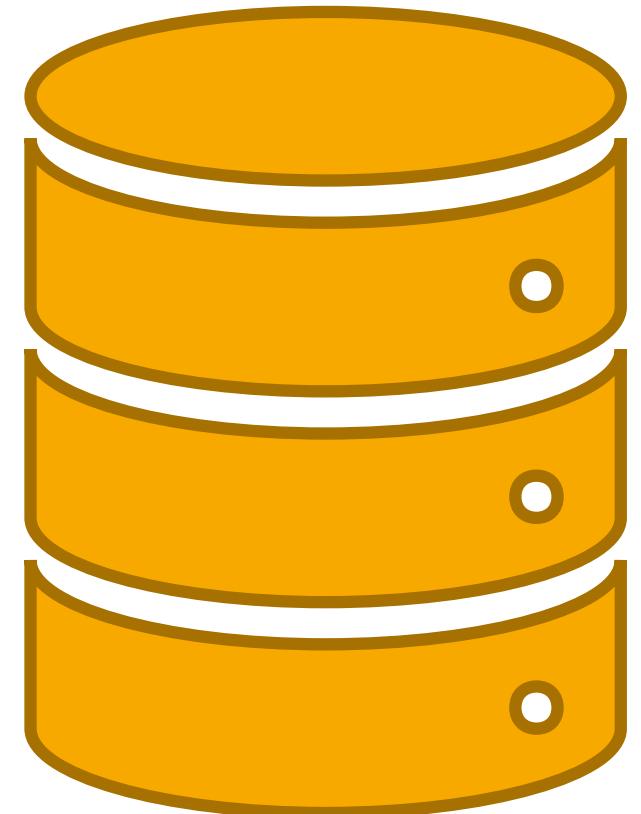


BANCO DE DADOS

Um banco de dados é uma coleção de dados relacionados

Onde **dados** são fatos conhecidos que podem ser registrados e que têm significado implícito.

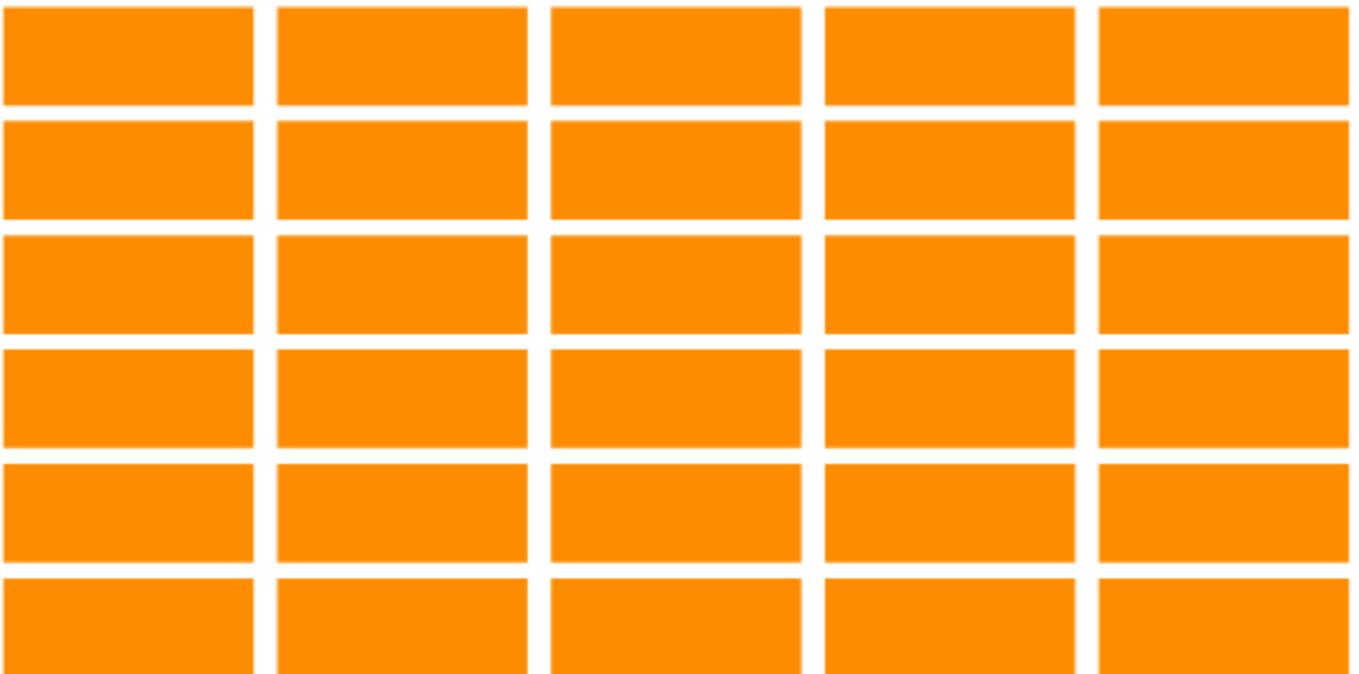
Por exemplo: nomes, números de telefone e endereços das pessoas que você conhece.



BANCO DE DADOS RELACIONAL



Um banco de dados relacional é uma coleção de **itens de dados organizados em um conjunto de tabelas com relacionamentos predefinidos** entre si.



Sistema de Gerenciamento de Banco de Dados (SGBD)



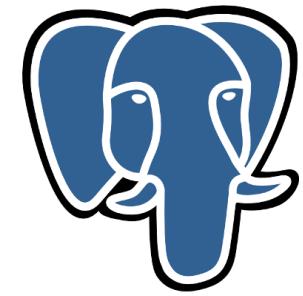
Um SGBD é Sistema computadorizado que permite aos usuários *criar* e manter um banco de dados.

ORACLE®
D A T A B A S E



Microsoft®
SQL Server®


MySQL™



Postgre**SQL**

STRUCTURED QUERY LANGUAGE (SQL)



Linguagem padrão para armazenar,
manipular e recuperar dados em
bancos de dados relacionais.



```
1 SELECT * FROM Estado;
```



O SURGIMENTO DOS SISTEMAS NOSQL

Com o avanço da tecnologia, o crescimento das Redes Sociais e sistemas cada vez mais interativos;





O SURGIMENTO DOS SISTEMAS NOSQL

Exigiu a criação de banco de dados, capazes de **armazenar** dados não tradicionais, diferentes de números e textos



O SURGIMENTO DOS SISTEMAS NOSQL



Sistemas de Armazenamento de Big Data ou Sistemas NoSQL, foram criados para gerenciar dados de aplicações de mídia social e análise de dados



Google
BIG TABLE



amazon
web services

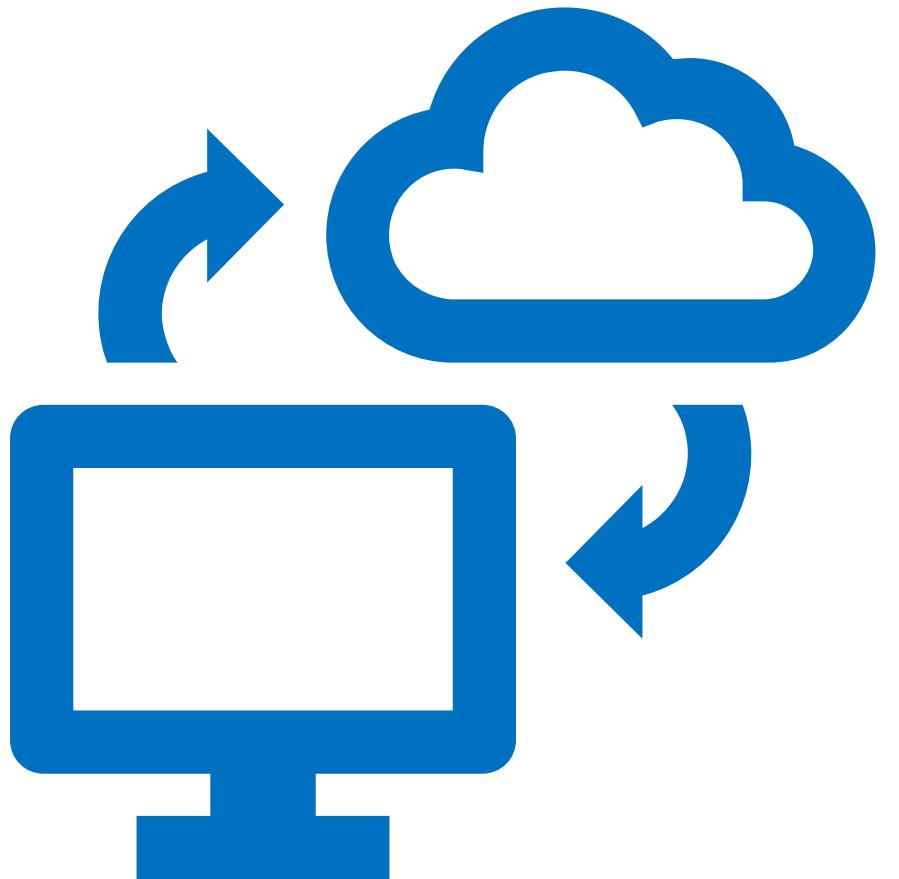
DynamoDB



COMPUTAÇÃO EM NUVEM (CLOUD COMPUTING)



Ambiente de computação em que o **software** e o **armazenamento** são fornecidos como um **serviço de internet** e acessado por meio de um navegador web.



COMPUTAÇÃO EM NUVEM (CLOUD COMPUTING)



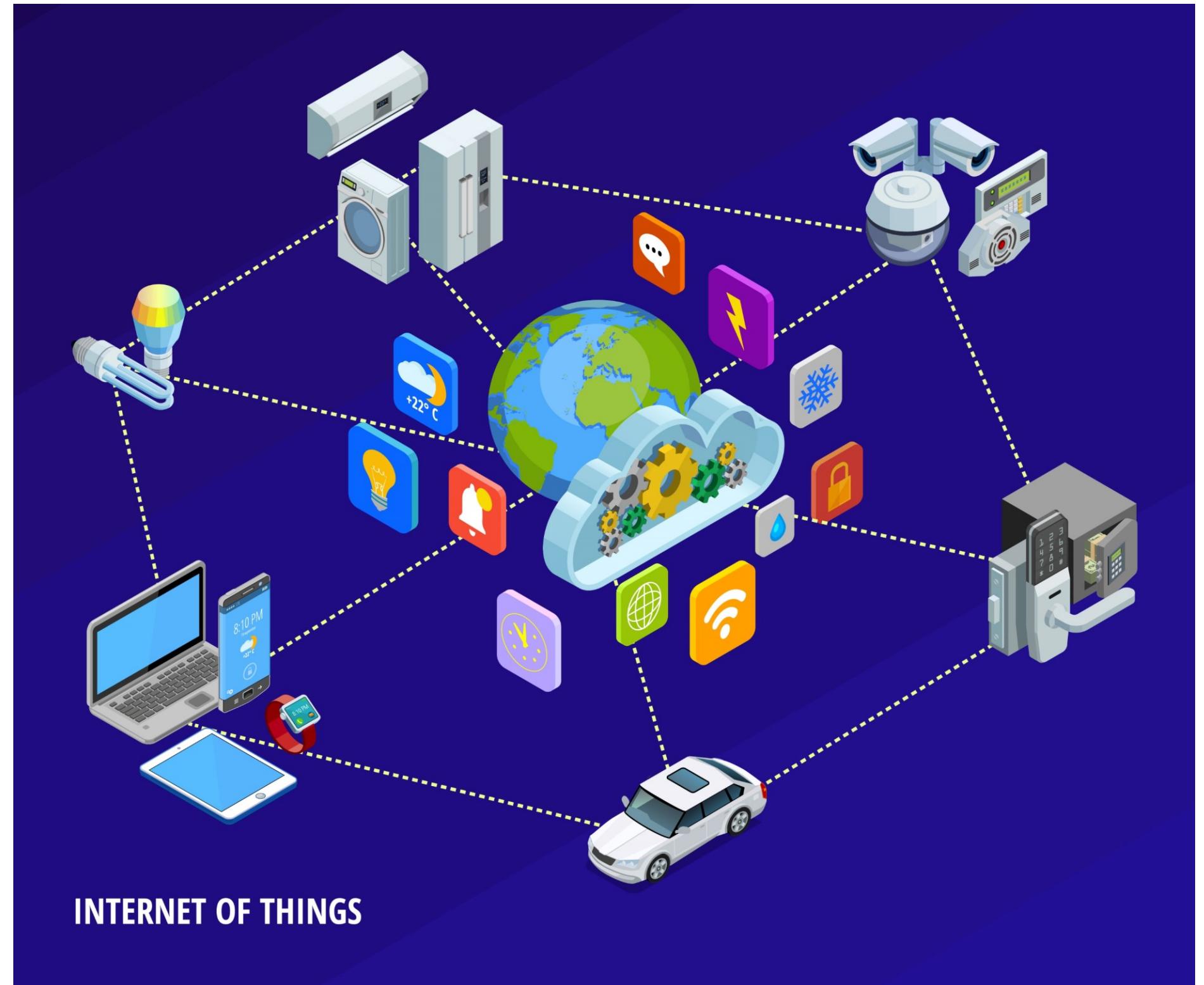
Plataformas destinadas à execução de aplicativos e serviços, baseada nos conceitos da computação em nuvem.





INTERNET DAS COISAS (IOT)

Rede de objetos físicos (*coisas*) com **sensores**, software e outras tecnologias com o objetivo de conectar e trocar dados com outros dispositivos e sistemas pela internet



IOT INDUSTRIAL



Aplicação da tecnologia IoT em ambientes industriais, especialmente no que diz respeito à instrumentação e controle de sensores e dispositivos que envolvem tecnologias de nuvem

- Fabricação inteligente;
- Ativos conectados e manutenção preventiva e preditiva;
- Redes de energia inteligentes;
- Cidades inteligentes;
- Logística conectada;
- Cadeias de suprimentos digitais inteligentes.

FERRAMENTAS DE VISUALIZAÇÃO DE INFORMAÇÃO



Ferramentas de análise de dados
que fornecem visualizações
interativas, dashboards e relatórios.



tableau
P A R T N E R



Sense™



Power BI

SOFTWARE ANACONDA



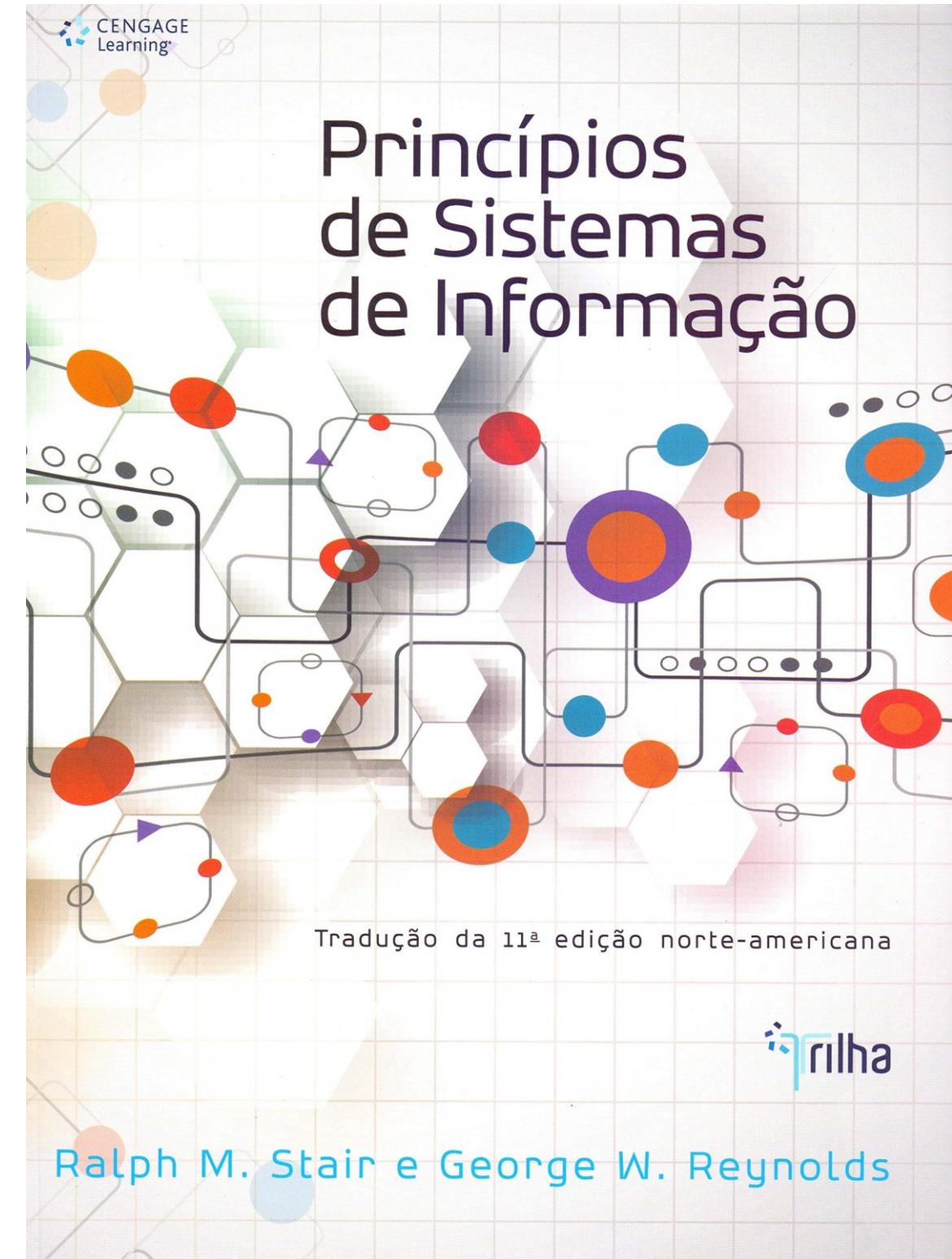
Distribuição das linguagens de programação Python e R para computação científica, que inclui pacotes para Ciência de Dados





BIBLIOGRAFIA PRINCIPAL

Princípios de sistemas de informação. Ralph Stair, George Reynolds. Cengage Learning. (2015). *Sistemas de Informações Empresariais - Parte 3, Cap. 10*





BIBLIOGRAFIA COMPLEMENTAR

Oracle. Big Data Definido

<https://www.oracle.com/br/big-data/what-is-big-data/>

ORACLE



BIBLIOGRAFIA COMPLEMENTAR

Oracle. O que é IOT?

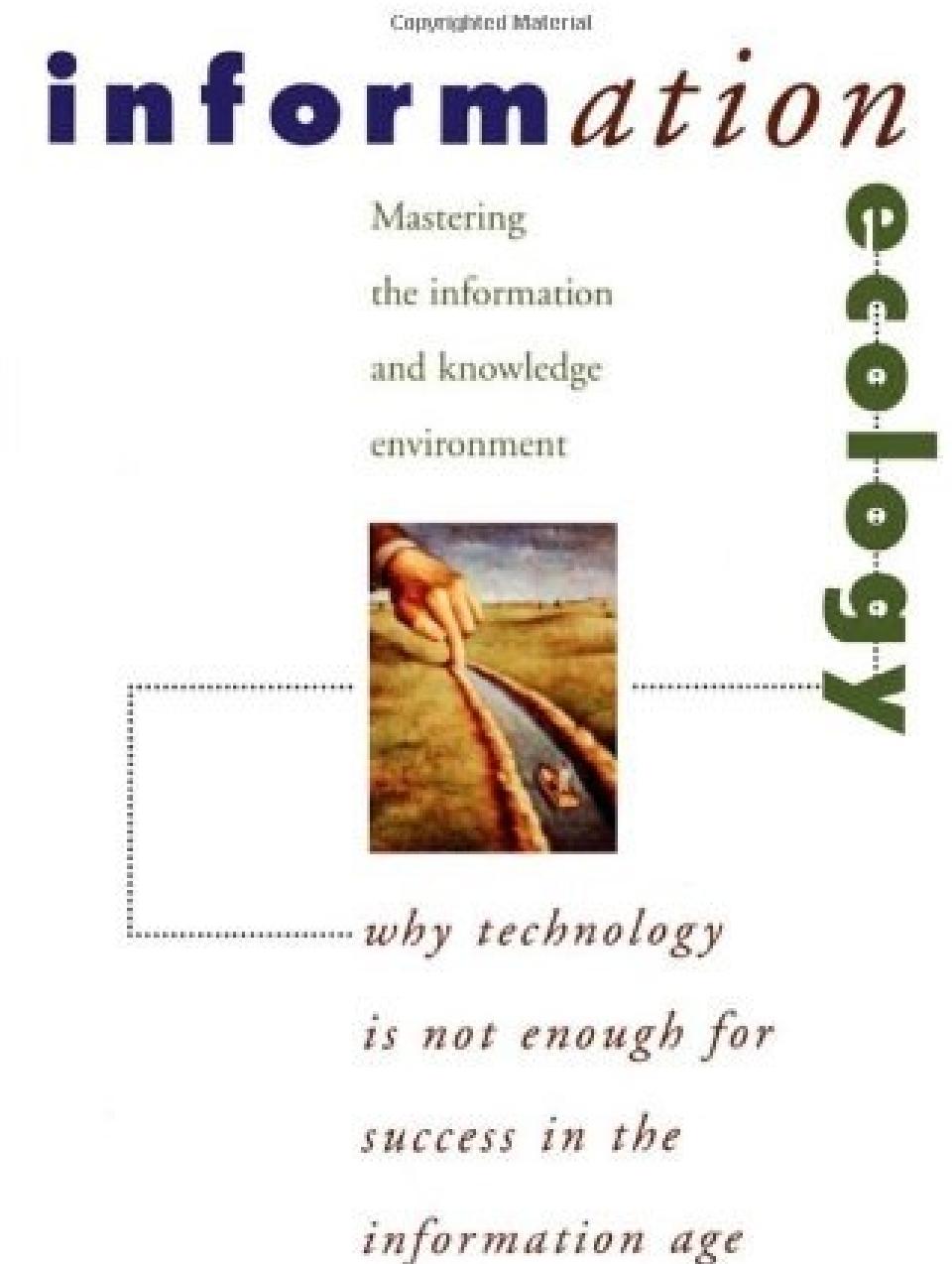
<https://www.oracle.com/br/internet-of-things/what-is-iot/>

ORACLE



BIBLIOGRAFIA COMPLEMENTAR

Davenport, Thomas H., and Laurence Prusak. *Information ecology: Mastering the information and knowledge environment*. Oxford University Press on Demand, 1997.

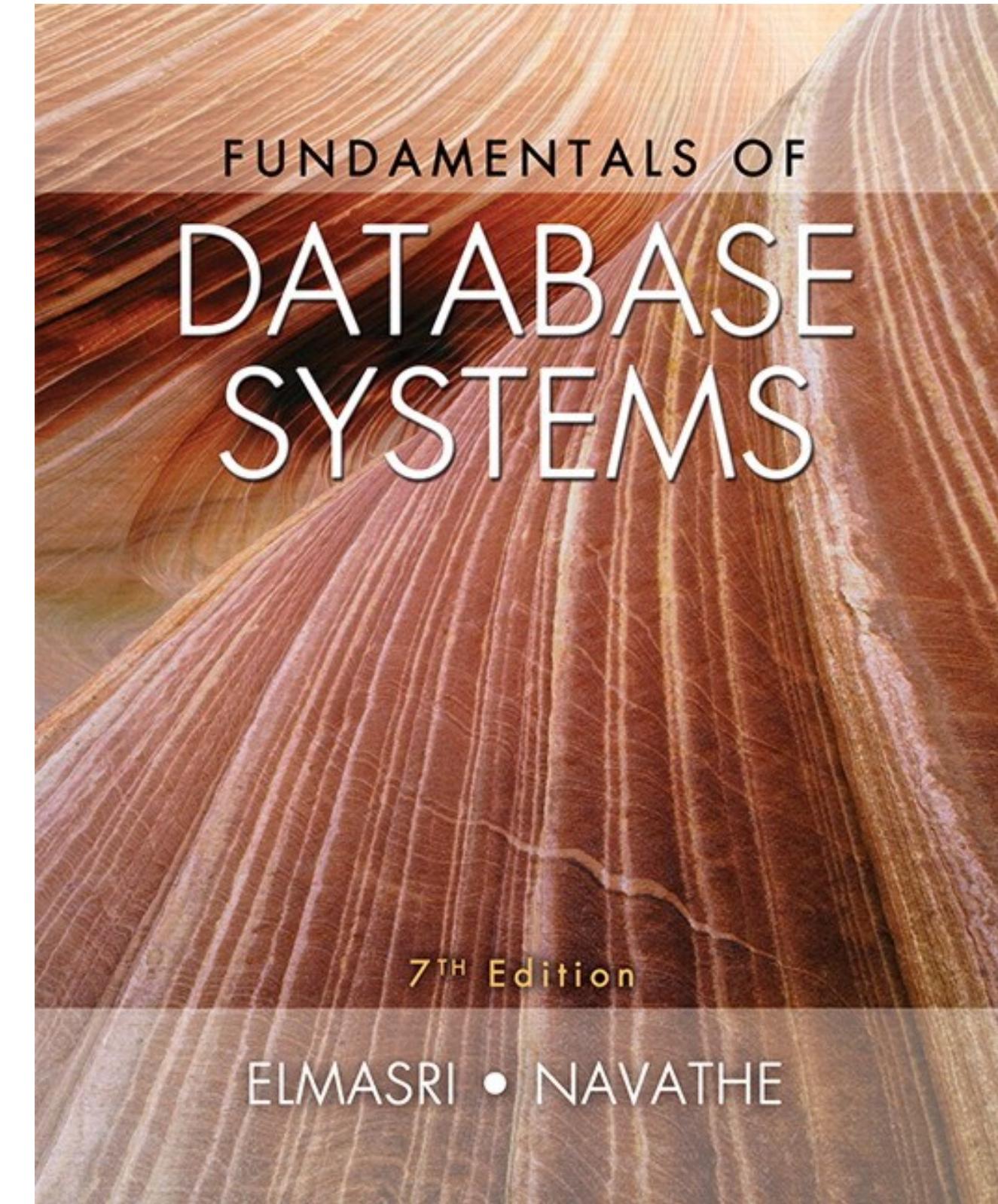


Thomas H. Davenport
Copyrighted Material



BIBLIOGRAFIA COMPLEMENTAR

R. Elmasri and S.
Navathe, Fundamentals
of database systems.
Pearson, 2017.





diaugustobarros@gmail.com



diegoaugustobarros.com



[@diegoaugustobarros](https://www.linkedin.com/in/diegoaugustobarros)



[@profdiegoaugusto](https://twitter.com/profdiegoaugusto)



Diego Augusto Barros
ENSINO & TECNOLOGIA

Prof. Diego Augusto

Diego Augusto Barros é bacharel em Sistemas de Informação pela Pontifícia Universidade Católica de Minas Gerais (2012) e mestre em Ciência da Computação pela Universidade Federal de Minas Gerais (2015). Sua pesquisa concentra-se nas áreas de Visualização de Dados e Interação Humano-computador, e investiga fatores cognitivos e perceptivos envolvidos na análise de grandes conjuntos de dados, que resultam em novos sistemas interativos para comunicação e análise visual. Seus principais interesses nas áreas são: visualização de informação, Visual Analytics, métodos de avaliação de interfaces, interação com sistemas, tecnologias web, sistemas de informação, engenharia de software e informática na educação.



LICENÇA

Esta obra está licenciada sob a Licença Internacional Creative Commons Attribution-NonCommercial-ShareAlike 4.0. Para ver uma cópia desta licença, visite:

<http://creativecommons.org/licenses/by-nc-sa/4.0/> ou envie uma carta para a Creative Commons, PO Box 1866, Mountain View, CA 94042, EUA.

