

# Cenário tecnológico atual

- Advento e enorme crescimento de plataformas on-line, apps de celulares, aplicativos em nuvem.
- Internet das coisas.
- Complexidade dos sistemas e das conexões. Necessidade de se automatizar tarefas.
- Dependência do mundo atual da tecnologia da informação.
- Todas as instituições ou empresas, de todos os segmentos, tornaram-se dependentes da tecnologia, de seus sistemas e bancos de dados.
- Tudo gera dados. A todo momento. Desde uma rede social como Instagram e Facebook, passando por ERPs, até sistemas complexos da área científica; o mundo está constantemente gerando dados.



- <https://www.worldometers.info/>

## O que é ciência de dados?

- Estudo dos dados para extrair insights significativos para as instituições, públicas ou privadas.
  - Possui abordagem multidisciplinar, convergindo conhecimentos de processos, tecnologias da informação, matemática e estatística.
- Possui como "matéria prima" os dados das empresas e instituições armazenados em servidores e bancos de dados.
- Explora, organiza e analisa o ativo mais precioso: Os dados, que, tratados da maneira correta, se





transformam em saber (conhecimento).

## Big Data – O que é

- Gigantesca massa de dados gerada através de diversas fontes e tipos de sistemas.
- Abrange vários tipos de arquivos: textos, áudios, imagens, bancos de dados estruturados e não estruturados.
- Abrange vários tipos de sistemas: Dados de sistemas operacionais, jogos, ERPs, CRMs, redes sociais, canais de streaming, servidores de arquivos, dados na nuvem, etc.
- **INFORMAÇÃO É PODER.** Para que dados se transformem em informação, é necessário ter ferramentas e metodologias que saibam trabalhar com essa grande massa de dados e extrair dela informações úteis e que possibilitem análise e tomadas de decisão.

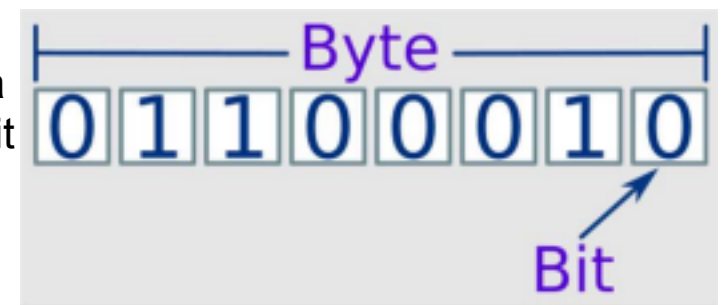


- Metodologia de racionalização da massa de dados. Estratificar e gerar informações compreensíveis de dados até então incompreensíveis.

## Medida e tamanho da informação • Exemplo: Para

representar a letra "a", precisamos de 1 byte: 01100001

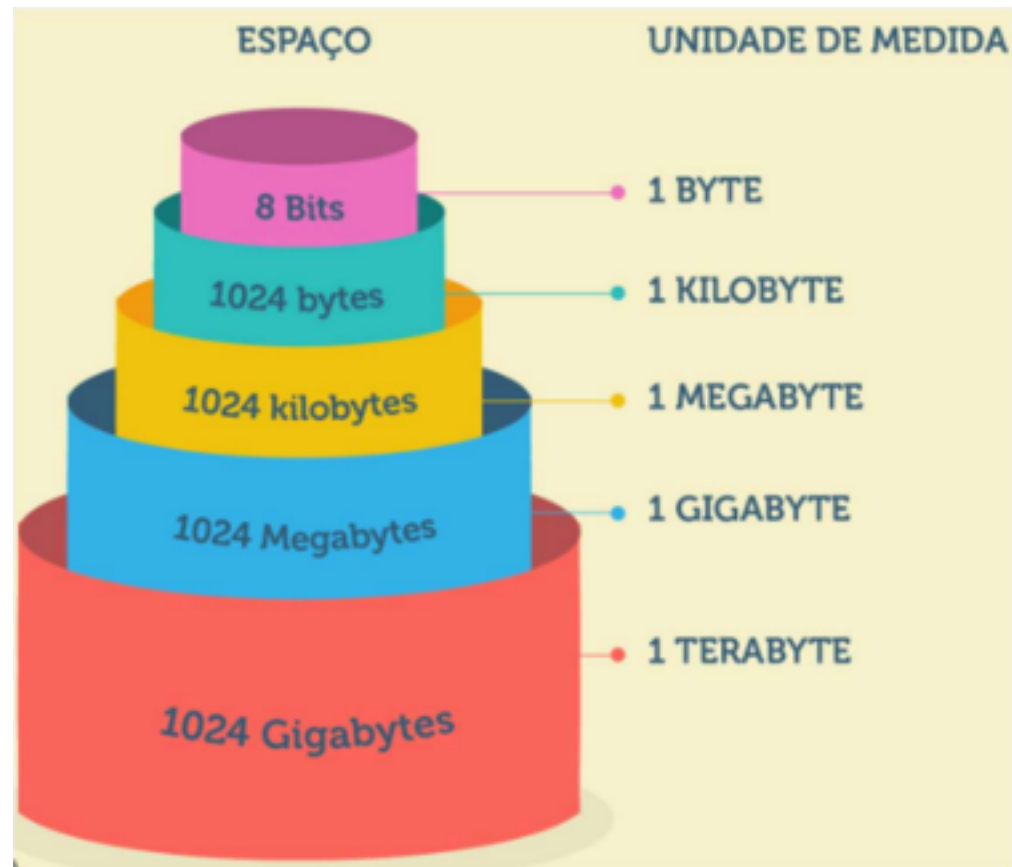
- Para entender a dimensão real da enorme quantidade de dados gerada atualmente é necessário conhecermos brevemente como funcionam as medidas de armazenamento em um sistema computacional.
- Os computadores entendem somente 0 e 1 (ausência ou presença de energia elétrica). Os computadores entendem apenas impulsos elétricos.
- A menor "porção" compreendida por um computador é chamada de binary digit (dígito binário). É o que chamamos de bit. Um bit pode assumir somente 0 ou 1.
- A combinação (junção) de 8 bits forma a menor unidade significativa da informação: o byte.



- Para armazenar o nome "Leticia" precisamos de 7 bytes, ou seja, 56 bits.

## Medida e tamanho da informação

- No contexto de armazenamento de dados dos computadores, a unidade de medida é em byte e seus respectivos múltiplos:



## Medida e tamanho da informação

- No contexto de armazenamento de dados dos computadores, a unidade de medida é em byte e seus respectivos múltiplos:



UNIDADE	Abreviação		Valor	Base Binária	Quantidade de Bytes
Bit			0 ou 1		
Byte			8 bits	$2^0$	1
Kilobyte	KByte	KB	1024 bytes	$2^{10}$	1.024
MegaByte	MByte	MB	1024 KB	$2^{20}$	1.048.576
Gigabyte	GByte	GB	1024 MB	$2^{30}$	1.073.741.824
Terabyte	TByte	TB	1024 GB	$2^{40}$	1.099.511.627.776
Petabyte	PByte	PB	1024 TB	$2^{50}$	1.125.899.906.842.620
Exabyte	EByte	EB	1024 PB	$2^{60}$	1.152.921.504.606.850.000
Zetabyte	ZByte	ZB	1024 EB	$2^{70}$	1.180.591.620.717.410.000.000
Yotabyte	YByte	YB	1024 ZB	$2^{80}$	1.208.925.819.614.630.000.000.000

## Medida e tamanho da informação

- Nos sistemas computacionais, os hardwares que armazenam a informação são chamados de "dispositivos de armazenamento" e podem ser de

diversos tipos, quantidades,  
tecnologias e valores:

- HD e SSD







- Storages e data centers

## Medida e tamanho da informação



- Data center Google:

<https://www.youtube.com/watch?v=zDAYZU4A3w0>

- Nos sistemas computacionais, os hardwares que armazenam a informação são chamados de "dispositivos de armazenamento" e podem ser de diversos tipos, quantidades, tecnologias e valores:

poder de processamento e armazenamento (storages profissionais). que podemos chamar de armazenamento distribuído.

## Medida e tamanho da informação

- Todo sistema que gera informação, armazena essas informações em algum dispositivo de armazenamento.

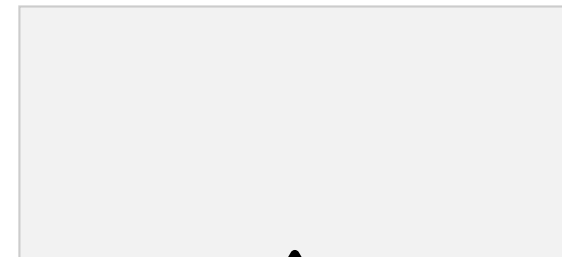
ERPs, CRMs e de BI? Seus softwares de logística?

- Esse "centro de armazenagem" pode ser também de diferentes tipos e objetivos.

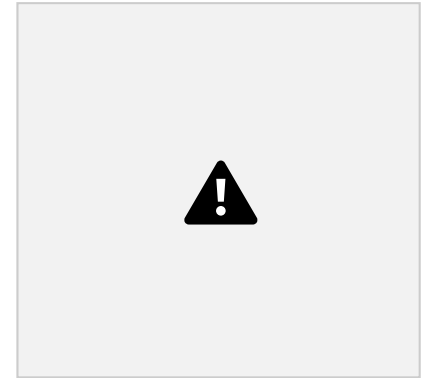
- Armazenamento local: Desde nosso computador até servidores de grande
- Armazenamento em nuvem: Dados armazenados em vários servidores espalhados por várias localidades do mundo. Em muitos casos assume o
- Onde armazenamos nossas fotos hoje? Onde as redes sociais armazenam nossas "curtidas", nossos comentários, as páginas e canais que seguimos? Onde estão armazenadas as séries e vídeos que assistimos?
- Onde as grandes empresas armazenam seus dados gerados de softwares analisadas dessa massa de dados?

**Medida e tamanho da informação** • Toda essa informação armazenada nos diversos tipos de dispositivos de armazenamento tem um enorme valor de análise.

- Exemplo: Uma grande loja online como a amazon.com.br: Imaginemos todas as compras feitas, a quantidade enorme de tipos de produtos, tipos de usuários, localidade dos usuários, volume dos produtos; imaginemos o sistema de logística, volume financeiro de compra e venda, parceiros de negócios envolvidos...

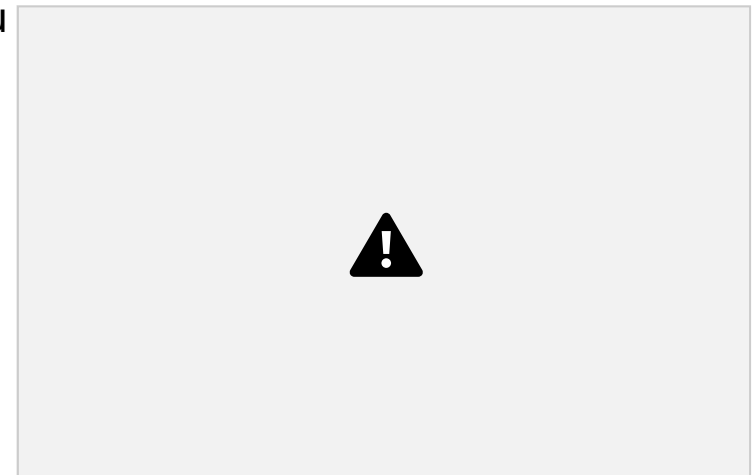


- Quantas informações são armazenadas e quantas informações podem ser
- Qual o perfil de um usuário que compra um Iphone? Quais produtos quem compra Iphone compra também? O sistema poderia gerar sugestões de compras? Qual localidade geográfica em um país mais adquire produtos?
- Quais os produtos mais vendidos? Por que? Existe alguma correlação entre o produto mais vendido e o perfil de consumo de uma região?
- Essa análise e essas perguntas podem ser respondidas pela análise de dados, uma das consequências do Big Data.



## Medida e tamanho da informação

- Pensemos também em um sistema de saúde de um país, ou em um sistema que colete informações sobre tipos de doenças no mundo todo.
- Imaginemos quantas informações poderíamos extrair e analisar. Informações essas que podem nos ajudar a entender padrões e tomar ações para melhorar ou



solucionar algum problema.

- De quais tipos de doença as pessoas no mundo todo mais morrem ou ficam com alguma sequela permanente? Existe alguma correlação desta doença com um padrão de vida, consumo de alimentos, região ou algum outro elemento que influencia nesses dados?
- Uma das metodologias para analisar toda essa massa de dados é através do Big Data e Data Analytics.

## Medida e tamanho da informação

- Resumo: Transformar uma enorme massa de dados em visões gráficas e numéricas para facilitar a compreensão e análise dos dados:



1. Coleta dos dados
2. Limpeza dos dados
3. Mineração dos dados
4. Análise de conteúdo
5. Visualização das informações



# Etapas de um projeto de análise de dados

- Trabalhar com toda essa massa de dados não é algo que se faz de forma aleatória ou de "qualquer jeito".
- Para bem implementar um projeto de análise de dados e/ou Big Data, existem processos (etapas) que devem ser respeitadas; caso contrário, a extração e geração das informações pode se tornar um fracasso.
- Um bom projeto de dados deve conter:

## 1º etapa: Coleta dos dados

- Para iniciar a coleta, deve ser definido um objetivo bem específico do
- A coleta dos dados é o primeiro passo de um projeto.

que se quer alcançar e de qual informação se necessita para realizar determinada análise.

- Exemplo: Uma empresa que queira obter informações do perfil de consumo de seus cliente online, que usem uma plataforma de compras.
- O objetivo é cruzar informações para conhecer o perfil de seus clientes.

- Informação que poderá ser coletada para realizar essa análise: Cliques em anúncios, dispositivos utilizados no momento da compra (computador, celular), localização do IP do cliente, histórico de compras, histórico de buscas, sistema operacional, etc.
- Neste exemplo, todas as informações acima fariam parte do conjunto de dados a serem coletados.

## 1º etapa: Coleta dos dados

- Outro exemplo: Uma grande empresa com várias filiais está tendo um aumento muito grande em suas despesas.
- Objetivo: Encontrar padrões ou desvios de padrões que justifiquem esse aumento das despesas.
- Dados a coletar: despesas fixas (conta de energia, aluguel da estrutura), despesas com prestadores de serviços, despesas com fretes, multas por contrato, enfim, todos os registros relacionados a despesas serão coletados.
- Após a coleta e junção desses dados, a empresa partirá para os próximos passos até montar uma estrutura visual dessas informações, a fim de tirar suas conclusões.

## 2º etapa: Limpeza dos dados

- A etapa de limpeza dos dados é uma das mais importantes. códigos de produtos, CPF, CNPJ, etc

e principalmente valores numéricos.

- É nessa etapa onde será feito um pré-processamento ou uma análise prévia dos dados, tendo por objetivo encontrar dados errados, anômalos, duplicados, nulos, fora de padrão, etc.
- Esses dados "sujos" devem ser removidos do conjunto ou em alguns casos corrigidos para um padrão aceitável.
- É de extrema importância realizar essa limpeza nos dados, pois ela irá prevenir que sejam geradas conclusões erradas sobre determinado assunto analisado.
- Dados "sujos" ou errados irão gerar por consequência uma visão de dados falsa. Isso seria algo péssimo para um projeto de análise de dados.
- Tipos de dados que geralmente precisam ser padronizados: datas,

**3° etapa: Mineração dos dados** • Nessa fase são usados métodos de machine learning, IA (inteligência

- Alguns exemplos dessas ferramentas: Tableau Public, Trifacta Wrangler.
- A etapa de mineração dos dados (Data Mining) é a etapa de processamento dos dados.
- Nesta etapa já conseguimos analisar as informações coletadas de uma forma global, já possibilitando uma tomada de decisões inicial.

- Nesta etapa são usadas ferramentas tecnológicas para sua execução. Somente com o olhar humano não seria possível analisar os dados em uma velocidade rápida (estamos falando de milhares de registros e dados coletados) e nem traçar padrões de forma puramente estatística.

artificial) e métodos estatísticos.

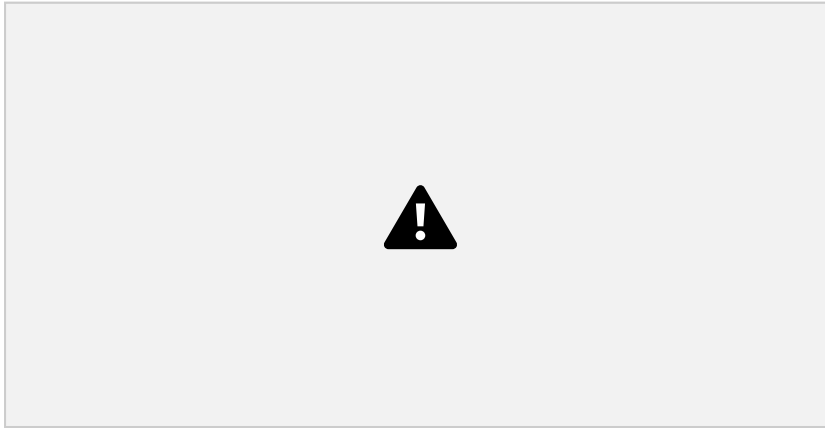
dentre várias outras disponíveis no mercado.

- Para saber mais: <https://www.tableau.com/pt-br/community/public>

## 4º etapa: Análise de conteúdo

- A etapa de análise de conteúdo deve ser realizada tendo o objetivo inicial como norte: Qual é o objetivo inicial do projeto? Ter insights de vendas? Verificar tendências futuras? Analisar decisões passadas?
- Nessa etapa a análise é feita através de gráficos ou infográficos. Não se trata apenas de pegar dados e transformá-los em gráficos. Se trata principalmente de transformá-los em elementos gráficos da forma mais adequada possível.

- Para cada situação existe um tipo de análise gráfica adequada.



## 4º etapa: Análise de conteúdo

- Tipos de análises:



## 4º etapa: Análise de conteúdo

- Análise descritiva: É o tipo de análise mais simples. Resumidamente consiste em analisar dados



já armazenados ou mesmo dados em tempo real visando uma análise do passado que então pode gerar uma mudança no futuro.

- Exemplo: Uma escola deseja verificar o nível de desistência de todos os cursos e entender quais cursos tem maior desistência.
- A escola precisará pegar os dados dos cursos e de suas desistências de todo o ano passado (ou outro período) e após analisar e cruzar esses dados (nome do curso, qual o tipo do curso, área do conhecimento, professor, etc) poderá visualizar quais são os cursos com maior desistência e todos os dados relacionados a ele.

**4º etapa: Análise de conteúdo** • A análise diagnóstica será responsável por ~~entender quais os motivos (causas)~~

- Análise diagnóstica: Como o próprio nome já diz, a análise diagnóstica permite avaliar, verificar, diagnosticar determinado padrão ou informações que motivaram determinado efeito, positivo ou negativo.
- Ela analisa o passado ou o presente para tirar conclusões (diagnósticos). • Está relacionada ao contexto causa --> efeito.
- Exemplo: Uma empresa sempre bate recorde de vendas nos meses de fevereiro e de setembro.

que levam a empresa a faturar mais nesses meses (clima, algum período festivo, algum investimento em marketing, etc).

**4º etapa: Análise de conteúdo** • São usadas técnicas de machine learning, IA e modelos estatísticos a fim de

- Análise preditiva: Nas análises descritivas e diagnósticas, a intenção é analisar o passado e o presente. A partir da análise preditiva o intuito é gerar uma análise do que possivelmente acontecerá no futuro.
- O objetivo é prever e se antecipar, para estar preparado para o que pode vir no futuro.

analisar tendências e comportamentos.

- Exemplo: Uma grande empresa de logística possui vários dados de manutenção de caminhões (km rodados, peças substituídas, trocas de pneus, paradas de veículos, etc).
- Com base nesses dados é possível prever e se prevenir de futuras paradas de frota. Exemplo: Determinado modelo de caminhão tende a se danificar aos 80.000 Km rodados. Com base nisso, a empresa pode se antecipar e aos 70 mil KM realizar uma manutenção preventiva para que a



parada não ocorra no meio de uma entrega.

**4º etapa: Análise de conteúdo** consegue nos fornecer qual seria o caminho ideal e mais rápido para chegarmos

- Análise prescritiva: A análise prescritiva também visa uma visão sobre o futuro, porém gerando prescrições, indicações e insights do que deve ser feito no futuro. • <http://ayata.com/energy/> - Empresa que desenvolve algoritmos prescritivos.

- Enquanto a análise descritiva visa fornecer uma visão sobre o que aconteceu e a análise preditiva ajuda a modelar e prever o que pode acontecer, a análise prescritiva busca determinar a melhor solução ou resultado entre as várias opções, dados os parâmetros conhecidos.

- Ela orienta quanto às consequências de decisões, antes delas serem tomadas, e recomenda caminhos possíveis a partir de cenários analisados.

- Um exemplo bem prático: O app Waze: Com base em várias informações (dados) tais como: horário, nível de tráfego, velocidade da via, dentre vários outros, ele

a um lugar. Temos a resposta pronta sem precisar analisar as diversas vias e rotas.

**4º etapa: Análise de conteúdo**



variando entre machine learning, estatística, teoria dos jogos, dentre outros.

- Análise prescritiva: Este tipo de análise envolve algoritmos e cálculos complexos,





**4º etapa: Análise de conteúdo** • Devemos ter em mente que esses 4 tipos de análise estão correlacionadas. Muitas vezes o resultado de uma determinada análise será o input de outra análise.

- Um bom cientista de dados deve conhecer essas análises, suas aplicações e objetivos.
- Alinhado com esse conhecimento dos tipos de análises, deve ter conhecimento também sobre o negócio (empresa ou instituição) que está analisando. Deve conhecer, através de pesquisas, entrevistas e com o apoio de outros consultores e da alta direção o objetivo e as expectativas da empresa.



- Deve ter visão holística.

## 5º etapa: Visualização das informações

- Nesta etapa o objetivo principal é desenvolver os gráficos ou infográficos finais com base em

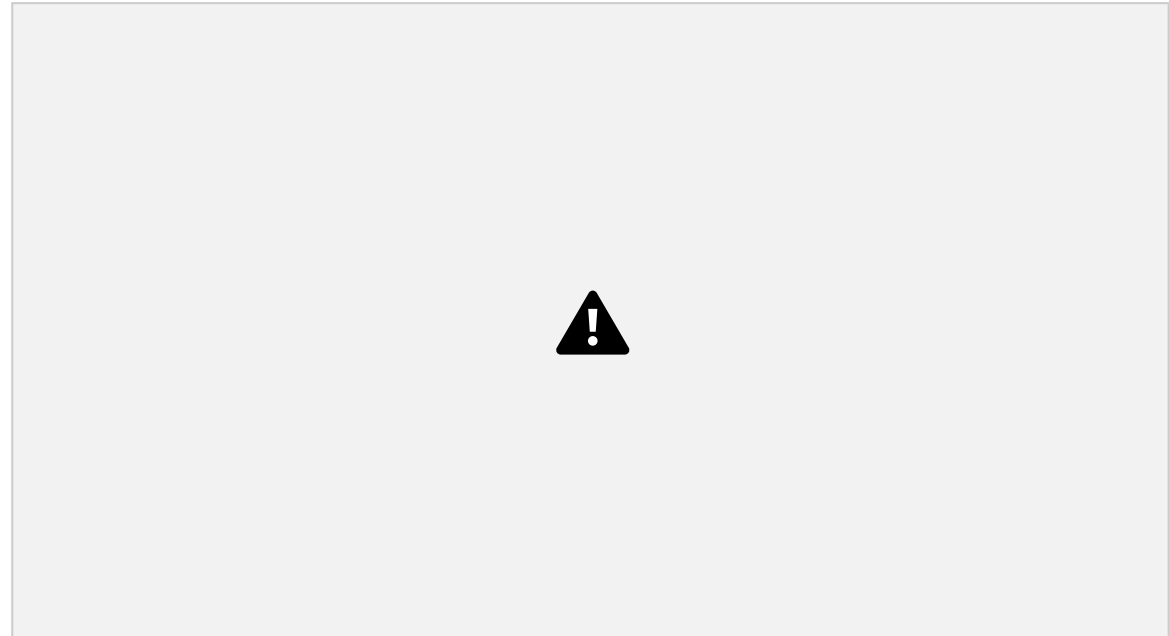


toda informação coletada e tratada nas etapas anteriores.

- O processo de geração dos gráficos visa facilitar a análise de determinado conteúdo, portanto a visualização deve ser customizada conforme o objetivo e conforme o público alvo que irá visualizar os resultados: diretores, gestores, colaboradores, etc.

- Exemplos de gráficos:

- 



<https://pt.venngage.com/blog/tipos-de-graficos/>

1 – Explique com suas próprias palavras o que é análise de dados. 2 – Qual a importância da

análise de dados e do Big Data no cenário atual das 3 – Por que se deve ter um objetivo

específico na implementação de um projeto de 4 – Existe diferença entre análise de dados e Big

Data? Se sim, qual?

## Atividade teórica

empresas?

Big Data ou de análise de dados?





## Banco de dados - Conceitos

Todos nós, de uma forma ou de outra, já interagimos com diversos bancos de dados durante nossa vida e continuamos interagindo no nosso dia-a-dia.

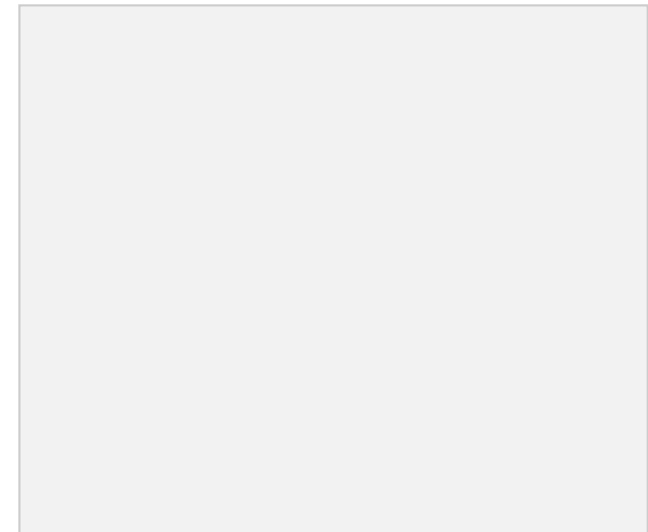
Pensemos por exemplo num extrato bancário, nas movimentações de nossa conta corrente. Todas as entradas e saídas são registradas e armazenadas para consulta e visualização.

Imaginemos o cadastro do SUS ou dos sistemas de convênios médicos, onde são registradas nossas informações de pacientes, nossos exames, médicos, clínicas e hospitais por onde passamos.

O mesmo ocorre nas redes sociais, onde se armazenam nossas fotos, heels, curtidas, likes e comentários. Tudo fica armazenado em um ou em vários bancos de dados.

## Banco de dados - Conceitos

Alguns exemplos:





## Banco de dados - Conceitos

Segundo Elmasri e Navathe\*, banco de dados podem ter os seguintes conceitos: • Um banco de dados é uma coleção logicamente coerente de dados com algum

\* RAMEZ ELMASRI; SHAMKANT NAVATHE. Sistemas de

banco de dados:

Pearson Education do Brasil Ltda. 2019

- Um banco de dados é uma coleção de dados relacionados.
- Um banco de dados representa algum aspecto do mundo real, às vezes chamado de minimundo ou de universo de discurso. As mudanças no minimundo são refletidas no banco de dados.

significado inerente. Uma variedade aleatória de dados não pode ser corretamente chamada de banco de dados.

- Um banco de dados é projetado, montado e preenchido com dados para uma finalidade específica. Ele tem um grupo intencionado de usuários e algumas aplicações previamente concebidas nas quais esses usuários estão interessados.

## Banco de dados SQL (estruturado)

De forma geral, um banco de dados possui a estrutura de campos e valores (colunas x linhas) em sua forma estruturada.

Cada campo guarda uma informação específica em uma determinada tabela.

Uma tabela pode ter vários campos, a depender da necessidade.





Exemplo: Uma tabela que guarda informações sobre alunos poderia ser da seguinte forma:



## Banco de dados SQL (estruturado)

Um banco de dados estruturado não tem uma única tabela. Um banco de dados bem construído terá várias tabelas.

Cada tabela é responsável por armazenar informações de um objeto (ou entidade).

Cada entidade possui seus próprios atributos. Vejamos o exemplo abaixo:



## Banco de dados SQL (estruturado)

No nosso exemplo a entidade é o "aluno" e seus atributos são cpf, nome, idade, sexo, telefone e cidade:





Como podemos perceber, o campo de cidade tem apenas um código. Esse código faz referência a uma outra tabela; à tabela de cidades.

Dessa forma, o modelo de banco de dados estruturado tem o conceito de "entidade-relacionamento". Vejamos:

## Banco de dados SQL (estruturado)

Tabela de cidades:

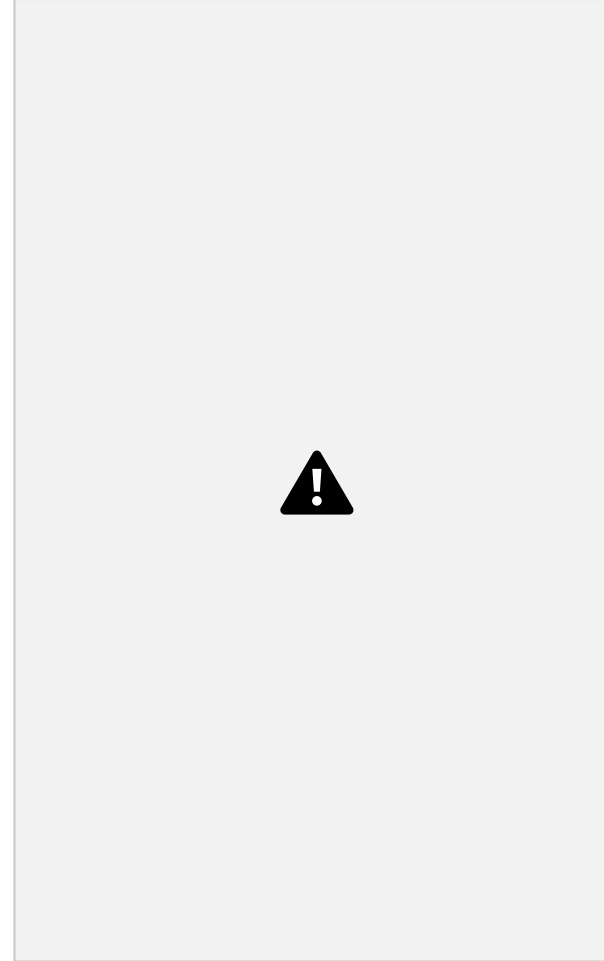
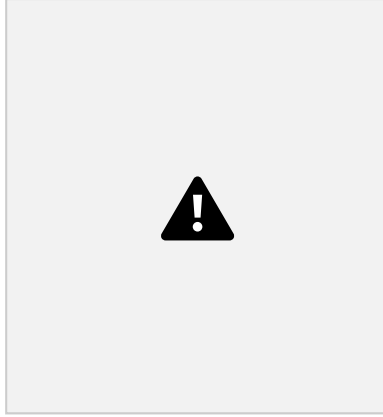


Tabela de alunos: Podemos perceber  
que cada código de cidade na tabela de alunos faz referência ao código da cidade na tabela de cidades:



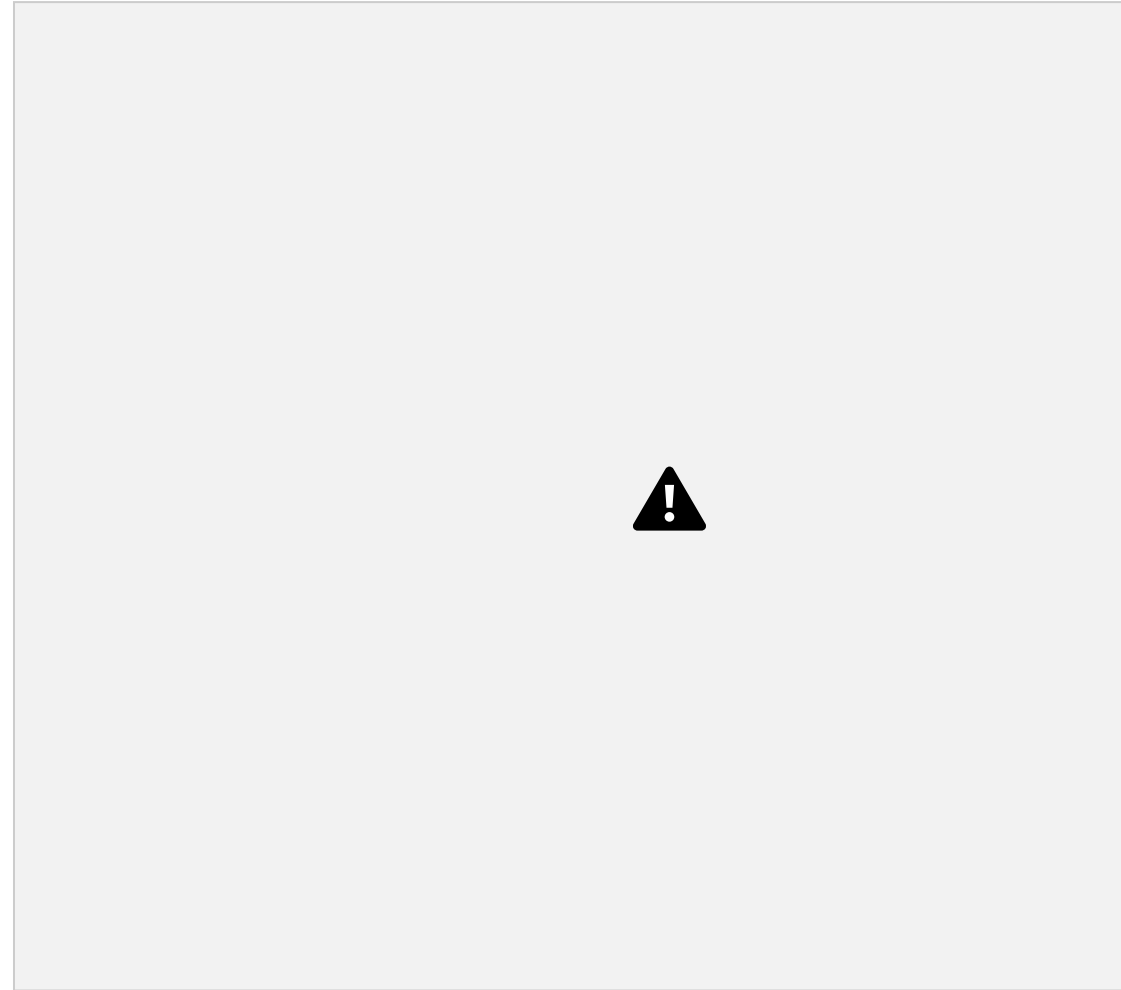
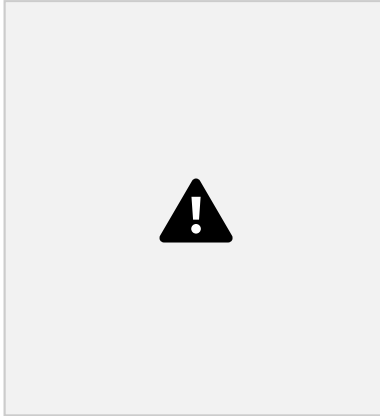
## Banco de dados SQL (estruturado)

Diagrama do banco de dados:



Banco de dados SQL (estruturado) Diagrama do banco

de dados. Vejamos o relacionamento desse banco de dados de exemplo:

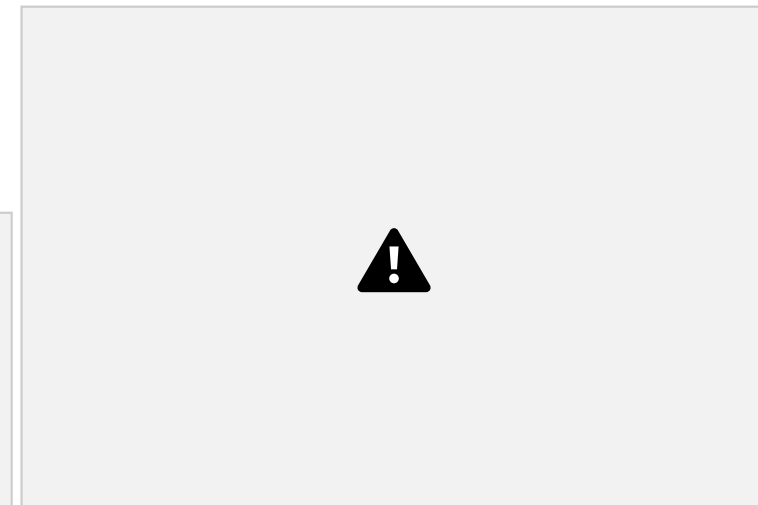
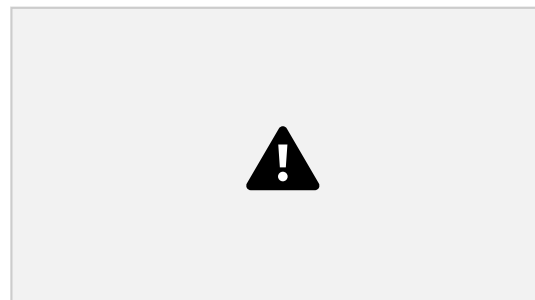
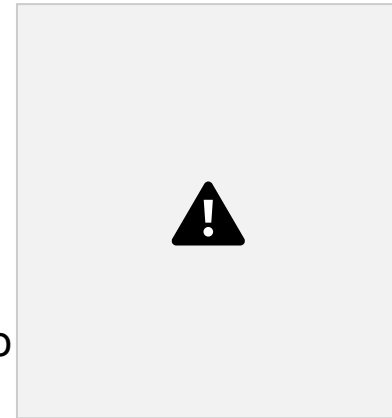


# Banco de dados SQL (estruturado)

É no banco de dados que as aplicações (softwares ERP, de logística, de compras online, etc) armazenam todos os cadastros e transações.

Desde uma matrícula em um curso até milhões de compras que são feitas diariamente por exemplo no site da Amazon.

Dessa forma, as diversas tabelas vão sendo populadas conforme a utilização dos aplicativos pelas empresas e usuários.



Por exemplo: Em um sistema onde são cadastradas as cidades, o banco de

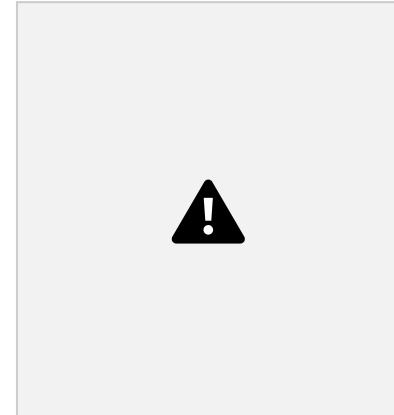
dados vai sendo construído em seu formato estruturado: tabelas. Com suas colunas e linhas:

## Banco de dados SQL (estruturado)

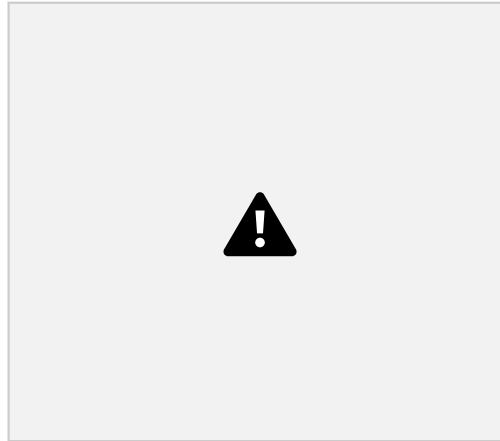
Em um banco de dados estrutura bem construído, todas as tabelas terão uma chave primária (que pode ser única ou composta por mais de um campo).

A chave primária é um valor que não pode ser repetido na tabela. O próprio mecanismo de banco de dados irá bloquear a inclusão de um valor chave primária caso o mesmo já exista no banco de dados.

Exemplo: Em uma tabela que recebe o cadastro de alunos, a chave primária seria o CPF e dessa forma o banco de dados iria garantir a integridade do cadastro, não permitindo que um usuário, por engano, cadastrasse um aluno em duplicidade:

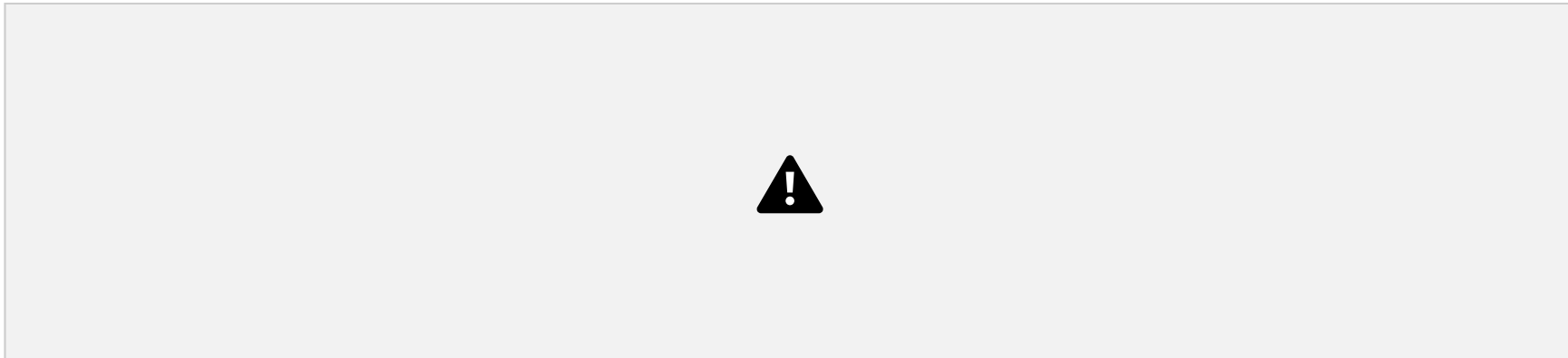






## Banco de dados SQL (estruturado)

Arquitetura:





Exemplo: Um sistema de compras online. A aplicação seria o site da Amazon com toda sua

## Banco de dados SQL (estruturado)

Arquitetura:

**Aplicação:** Aplicativo que será a interface entre o usuário final e o banco de dados. interface e produtos disponíveis.

**SGBD:** Sistema gerenciador de banco de dados. É um software que possibilita a gestão e manuseio dos dados. É neste software que o responsável pela gestão ou visualização dos dados pode trabalhar. É onde se coloca a "mão na massa".



**Banco de dados:** É onde os dados ficam armazenados. Sistema complexo e robusto para armazenar grandes quantidades de registros. Possui arquitetura completa e tratativas relacionadas a desempenho, gestão e volumetria dos dados.

consulta estruturada (Structured query language – SQL).

## Banco de dados SQL (estruturado)

Um analista de dados (ou qualquer outro profissional que queira manusear dados) deve saber extrair informações de bancos de dados.

Essas informações podem ser extraídas através do SGBD, usando uma linguagem de

É uma linguagem extremamente conhecida no mundo de bancos de dados. Várias empresas e bancos de dados a utilizam como ferramenta.

Alguns exemplos de softwares/aplicações de bancos de dados: MySQL (Oracle), SQL Server (Microsoft), PostgreSQL, DB2 (IBM), dentre outros.



