

# Métodos Computacionais Aplicados (CEC0031)

## Curso Introductório de R

Prof.: Dr. José Weligton Félix Gomes

Universidade Federal do Ceará - UFC  
Campus Avançado de Sobral

13 de março de 2023

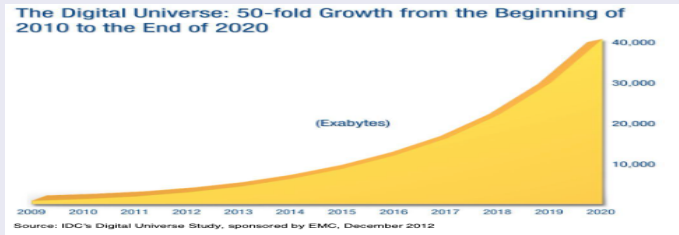
- Em reportagem veiculada no site do **OPEN KNOWLEDGE BRASIL**, em 29 de Setembro de 2017, cujo título "**O que faremos com os 40 trilhões de gigabytes de dados disponíveis em 2020?**", Thiago Ávila destaca que o volume de dados tem crescido exponencialmente.
- O autor apresenta o estudo "*A Universe of Opportunities and Challenges*", no qual, aponta que de 2006 a 2010, o volume de dados digitais gerados cresceu de **166 Exabytes** para **988 Exabytes**.
- Existe a perspectiva que o volume de dados alcance a casa dos **40.000 Exabytes**, ou **40 Zettabytes** (ou **40 trilhões de Gigabytes**) em 2020.
- Segundo o site **Infor Channel** este número seria ainda maior, **44 Zettabytes**, como prevê o estudo do Instituto *for the Future* (ITFF).

- Em reportagem veiculada no site do **OPEN KNOWLEDGE BRASIL**, em 29 de Setembro de 2017, cujo título "**O que faremos com os 40 trilhões de gigabytes de dados disponíveis em 2020?**", Thiago Ávila destaca que o volume de dados tem crescido exponencialmente.
- O autor apresenta o estudo "*A Universe of Opportunities and Challenges*", no qual, aponta que de 2006 a 2010, o volume de dados digitais gerados cresceu de **166 Exabytes** para **988 Exabytes**.
- Existe a perspectiva que o volume de dados alcance a casa dos **40.000 Exabytes**, ou **40 Zettabytes** (ou **40 trilhões de Gigabytes**) em 2020.
- Segundo o site **Infor Channel** este número seria ainda maior, **44 Zettabytes**, como prevê o estudo do Instituto *for the Future* (ITFF).

- Em reportagem veiculada no site do **OPEN KNOWLEDGE BRASIL**, em 29 de Setembro de 2017, cujo título "**O que faremos com os 40 trilhões de gigabytes de dados disponíveis em 2020?**", Thiago Ávila destaca que o volume de dados tem crescido exponencialmente.
- O autor apresenta o estudo "*A Universe of Opportunities and Challenges*", no qual, aponta que de 2006 a 2010, o volume de dados digitais gerados cresceu de **166 Exabytes** para **988 Exabytes**.
- Existe a perspectiva que o volume de dados alcance a casa dos **40.000 Exabytes**, ou **40 Zettabytes** (ou **40 trilhões de Gigabytes**) em 2020.
- Segundo o site **Infor Channel** este número seria ainda maior, **44 Zettabytes**, como prevê o estudo do Instituto *for the Future* (ITFF).

- Em reportagem veiculada no site do **OPEN KNOWLEDGE BRASIL**, em 29 de Setembro de 2017, cujo título "**O que faremos com os 40 trilhões de gigabytes de dados disponíveis em 2020?**", Thiago Ávila destaca que o volume de dados tem crescido exponencialmente.
- O autor apresenta o estudo "*A Universe of Opportunities and Challenges*", no qual, aponta que de 2006 a 2010, o volume de dados digitais gerados cresceu de **166 Exabytes** para **988 Exabytes**.
- Existe a perspectiva que o volume de dados alcance a casa dos **40.000 Exabytes**, ou **40 Zettabytes** (ou **40 trilhões de Gigabytes**) em 2020.
- Segundo o site **Infor Channel** este número seria ainda maior, **44 Zettabytes**, como prevê o estudo do Instituto *for the Future* (ITFF).

# Motivação



## Prefixos Métricos e Unidades do Sistema Internacional

Prefixo (Símbolo)	Em Potência de 2	Em Potência de 10	Representação numérica
Yottabyte (YB)	$2^{80}$	$10^{24}$	1 setilhão
Zettabyte (ZB)	$2^{70}$	$10^{21}$	1 sextilhão
Exabyte (EB)	$2^{60}$	$10^{18}$	1 quintilhão
Petabyte (PB)	$2^{50}$	$10^{15}$	1 quadrilhão
Terabyte (TB)	$2^{40}$	$10^{12}$	1 trilhão
Gigabyte (GB)	$2^{30}$	$10^9$	1 bilhão
Megabyte (MB)	$2^{20}$	$10^6$	1 milhão
Quilobyte (KB)	$2^{10}$	$10^3$	1 mil
Byte (B)	$2^0$	$10^0$	1 unidade

- Thiago Ávila, em artigo do dia 23 de outubro de 2017, com título "Entendendo a atuação do W3C para a Web dos Dados (fala do W3C Web Activity)", também afirmou que "o crescimento da oferta de dados global vem trazendo desafios interessantes e, entre os caminhos para o melhor aproveitamento do potencial destes dados, a "Web dos Dados" vem se consolidando como o novo paradigma para os próximos ano".
- "Novas abordagens em torno dos dados foram sendo desenvolvidas ao longo dos anos e atualmente, busca-se o estabelecimento de um conceito de dado que possa ser amplamente utilizado sem restrições de uso e aplicações, de tal maneira que o ciclo de produção de conhecimento possa ser mais rico e aprimorado. (Bandeira *et. al*, 2014).
- Sendo assim, "os dados abertos permitem que pessoas e organizações utilizem informações públicas livremente para gerar aplicativos, fazer análises ou mesmo produtos comercializáveis".



- Thiago Ávila, em artigo do dia 23 de outubro de 2017, com título "Entendendo a atuação do W3C para a Web dos Dados (fala do W3C Web Activity)", também afirmou que "o crescimento da oferta de dados global vem trazendo desafios interessantes e, entre os caminhos para o melhor aproveitamento do potencial destes dados, a "Web dos Dados" vem se consolidando como o novo paradigma para os próximos ano".
- "Novas abordagens em torno dos dados foram sendo desenvolvidas ao longo dos anos e atualmente, busca-se o estabelecimento de um conceito de dado que possa ser amplamente utilizado sem restrições de uso e aplicações, de tal maneira que o ciclo de produção de conhecimento possa ser mais rico e aprimorado. (Bandeira *et. al*, 2014).
- Sendo assim, "os dados abertos permitem que pessoas e organizações utilizem informações públicas livremente para gerar aplicativos, fazer análises ou mesmo produtos comercializáveis".

- Thiago Ávila, em artigo do dia 23 de outubro de 2017, com título "Entendendo a atuação do W3C para a Web dos Dados (fala do W3C Web Activity)", também afirmou que "o crescimento da oferta de dados global vem trazendo desafios interessantes e, entre os caminhos para o melhor aproveitamento do potencial destes dados, a "Web dos Dados" vem se consolidando como o novo paradigma para os próximos ano".
- "Novas abordagens em torno dos dados foram sendo desenvolvidas ao longo dos anos e atualmente, busca-se o estabelecimento de um conceito de dado que possa ser amplamente utilizado sem restrições de uso e aplicações, de tal maneira que o ciclo de produção de conhecimento possa ser mais rico e aprimorado. (Bandeira *et. al*, 2014).
- Sendo assim, "os dados abertos permitem que pessoas e organizações utilizem informações públicas livremente para gerar aplicativos, fazer análises ou mesmo produtos comercializáveis".

# Introdução à Ciência de Dados

- Conforme **Oliveira, Guerra, McDonnell (2018)**:

Trata-se de um termo que designa uma área de conhecimento voltada para o **estudo** e a **análise de dados**, onde busca-se extrair conhecimento e criar novas informações.

- É uma atividade interdisciplinar que concilia, principalmente, duas grandes áreas: Ciência da Computação e Estatística.
- A Ciência de Dados vem sendo aplicada como apoio em diferentes outras áreas de conhecimento, tais como: Medicina, Biologia, Economia, Comunicação, Ciências Políticas, etc.
- Apesar de não ser uma área nova, o tema vem se popularizando cada vez mais, graças à explosão na produção de dados e crescente dependência dos dados para a tomada de decisão.

# Introdução à Ciência de Dados

- Conforme **Oliveira, Guerra, McDonnell (2018)**:

Trata-se de um termo que designa uma área de conhecimento voltada para o **estudo** e a **análise de dados**, onde busca-se extrair conhecimento e criar novas informações.

- É uma atividade interdisciplinar que concilia, principalmente, duas grandes áreas: Ciência da Computação e Estatística.
- A Ciência de Dados vem sendo aplicada como apoio em diferentes outras áreas de conhecimento, tais como: Medicina, Biologia, Economia, Comunicação, Ciências Políticas, etc.
- Apesar de não ser uma área nova, o tema vem se popularizando cada vez mais, graças à explosão na produção de dados e crescente dependência dos dados para a tomada de decisão.

# Introdução à Ciência de Dados

- Conforme **Oliveira, Guerra, McDonnell (2018)**:

Trata-se de um termo que designa uma área de conhecimento voltada para o **estudo** e a **análise de dados**, onde busca-se extrair conhecimento e criar novas informações.

- É uma atividade interdisciplinar que concilia, principalmente, duas grandes áreas: Ciência da Computação e Estatística.
- A Ciência de Dados vem sendo aplicada como apoio em diferentes outras áreas de conhecimento, tais como: Medicina, Biologia, Economia, Comunicação, Ciências Políticas, etc.
- Apesar de não ser uma área nova, o tema vem se popularizando cada vez mais, graças à explosão na produção de dados e crescente dependência dos dados para a tomada de decisão.

# Introdução à Ciência de Dados

- Conforme **Oliveira, Guerra, McDonnell (2018)**:

Trata-se de um termo que designa uma área de conhecimento voltada para o **estudo** e a **análise de dados**, onde busca-se extrair conhecimento e criar novas informações.

- É uma atividade interdisciplinar que concilia, principalmente, duas grandes áreas: Ciência da Computação e Estatística.
- A Ciência de Dados vem sendo aplicada como apoio em diferentes outras áreas de conhecimento, tais como: Medicina, Biologia, Economia, Comunicação, Ciências Políticas, etc.
- Apesar de não ser uma área nova, o tema vem se popularizando cada vez mais, graças à explosão na produção de dados e crescente dependência dos dados para a tomada de decisão.

# Introdução à Ciência de Dados

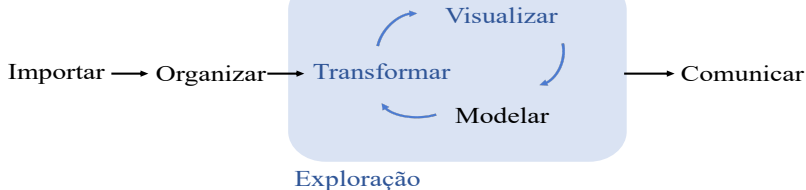
- Conforme **Oliveira, Guerra, McDonnell (2018)**:

Trata-se de um termo que designa uma área de conhecimento voltada para o **estudo** e a **análise de dados**, onde busca-se extrair conhecimento e criar novas informações.

- É uma atividade interdisciplinar que concilia, principalmente, duas grandes áreas: Ciência da Computação e Estatística.
- A Ciência de Dados vem sendo aplicada como apoio em diferentes outras áreas de conhecimento, tais como: Medicina, Biologia, Economia, Comunicação, Ciências Políticas, etc.
- Apesar de não ser uma área nova, o tema vem se popularizando cada vez mais, graças à explosão na produção de dados e crescente dependência dos dados para a tomada de decisão.

# Workflow da Ciência de Dados

- Não existe apenas uma forma de estruturar e aplicar os conhecimentos da Ciência de Dados. Vejamos, por exemplo, o *workflow* apresentado no livro **R for Data Science** (WICKHAM e GROLEMUND, 2017).





- Esse **workflow** propõe basicamente os seguintes passos:

- 1 Carregar os dados;
- 2 Limpar os dados;
- 3 Transformar, visualizar e modelar (fase exploratória);
- 4 Comunicar o resultado.

# Workflow da Ciência de Dados

- Esse **workflow** propõe basicamente os seguintes passos:
  - 1 Carregar os dados;
  - 2 Limpar os dados;
  - 3 Transformar, visualizar e modelar (fase exploratória);
  - 4 Comunicar o resultado.

# Workflow da Ciência de Dados

- Esse **workflow** propõe basicamente os seguintes passos:
  - 1 Carregar os dados;
  - 2 Limpar os dados;
  - 3 Transformar, visualizar e modelar (fase exploratória);
  - 4 Comunicar o resultado.

- Esse **workflow** propõe basicamente os seguintes passos:
  - 1 Carregar os dados;
  - 2 Limpar os dados;
  - 3 Transformar, visualizar e modelar (fase exploratória);
  - 4 Comunicar o resultado.

- Esse **workflow** propõe basicamente os seguintes passos:
  - ① Carregar os dados;
  - ② Limpar os dados;
  - ③ Transformar, visualizar e modelar (fase exploratória);
  - ④ Comunicar o resultado.

# Linguagens para Ciência de dados

- Existem diversas linguagens ou ferramentas para exploração de dados: R, Python, SAS, SQL, Matlab, Stata, Power BI, etc...
- Cabe ao cientista de dados avaliar qual é a ferramenta mais adequada para alcançar seus objetivos.

- Existem diversas linguagens ou ferramentas para exploração de dados: R, Python, SAS, SQL, Matlab, Stata, Power BI, etc...
- Cabe ao cientista de dados avaliar qual é a ferramenta mais adequada para alcançar seus objetivos.