

D1INT – Introdução à Ciência de Dados

2021.1

Aula 01.2

Introdução – Definindo Ciência de Dados



Prof. Everton Silva
everton.silva@ifsp.edu.br



Aula de Hoje

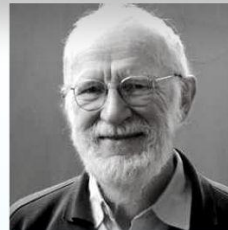
- Uma breve história
- O que é Ciência de Dados?
- Os dados na Ciência de Dados.

Uma breve história

HISTORY OF DATA SCIENCE



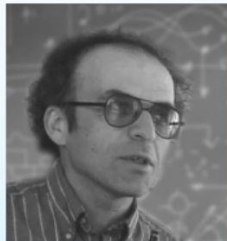
1962, John W. Tukey writes in "The Future of Data Analysis"



1974, Peter Naur publishes the Concise Survey of Computer Methods



1977, the International Association for Statistical Computing (IASC) was founded



1989, Gregory Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop

BusinessWeek

1994, BusinessWeek published a cover story on "Database Marketing"



1997, during his inaugural lecture as the H. C. Carver Chair in Statistics at the University of Michigan, Jeff Wu called for statistics to be renamed "data science" and statisticians to be renamed "data scientists".

1996, in the occasion of the conference of International Federation of Classification Societies (IFCS), for the first time, the term "data science" is included in the title of the conference



Infographic by @ingliguori



O que é Ciência de Dados?

Algumas definições

A ciência de dados se preocupa em analisar dados e extrair conhecimento útil deles. A construção de modelos preditivos geralmente é a atividade mais importante para um cientista de dados ¹.

A ciência de dados se preocupa em analisar Big Data para extrair correlações com estimativas de probabilidade e erro ².

A ciência de dados é uma disciplina emergente que se baseia no conhecimento da metodologia estatística e da ciência da computação para criar previsões e percepções impactantes para uma ampla gama de campos acadêmicos tradicionais ³.

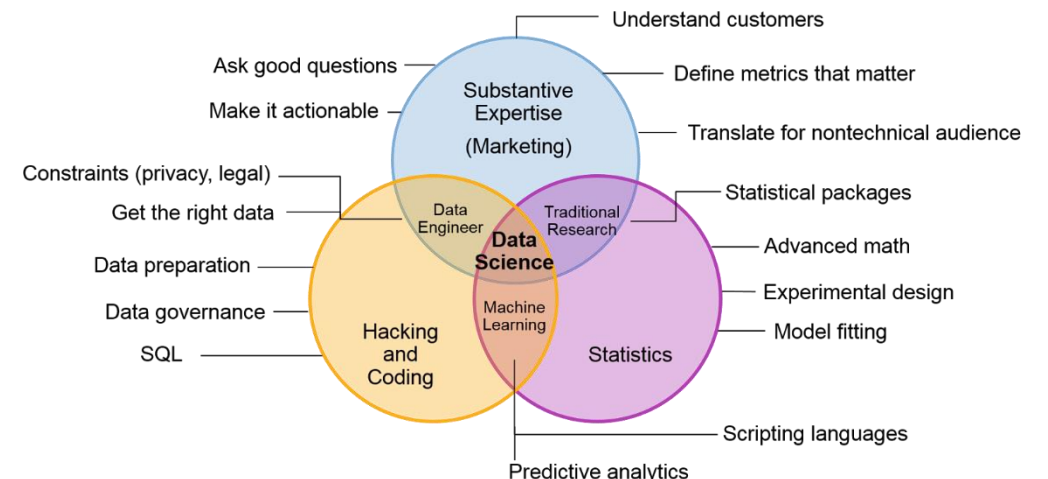
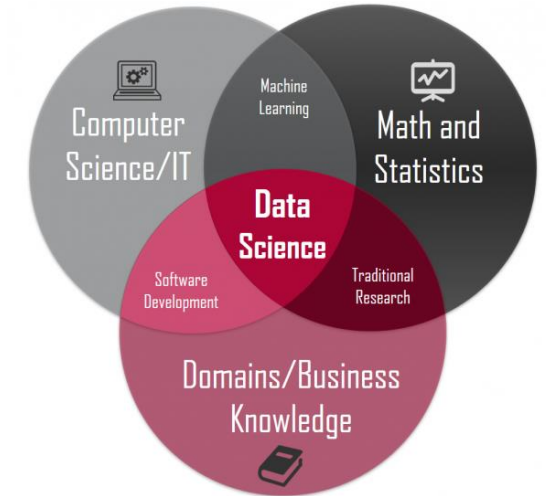
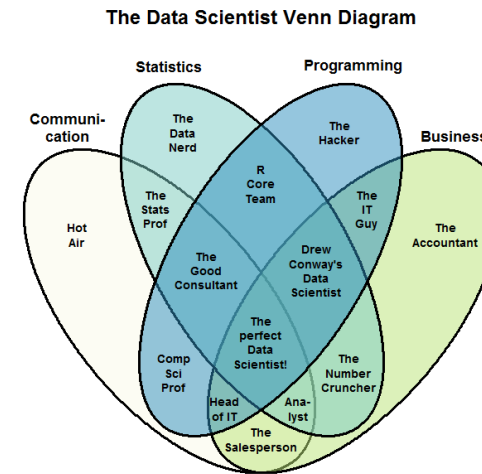
1. <https://www.kdnuggets.com/tag/data-science>

2. https://www.researchgate.net/publication/285586313_Understanding_Data_Science_An_Emerging_Discipline_for_Data-Intensive_Discovery

3. <https://datascience.harvard.edu>

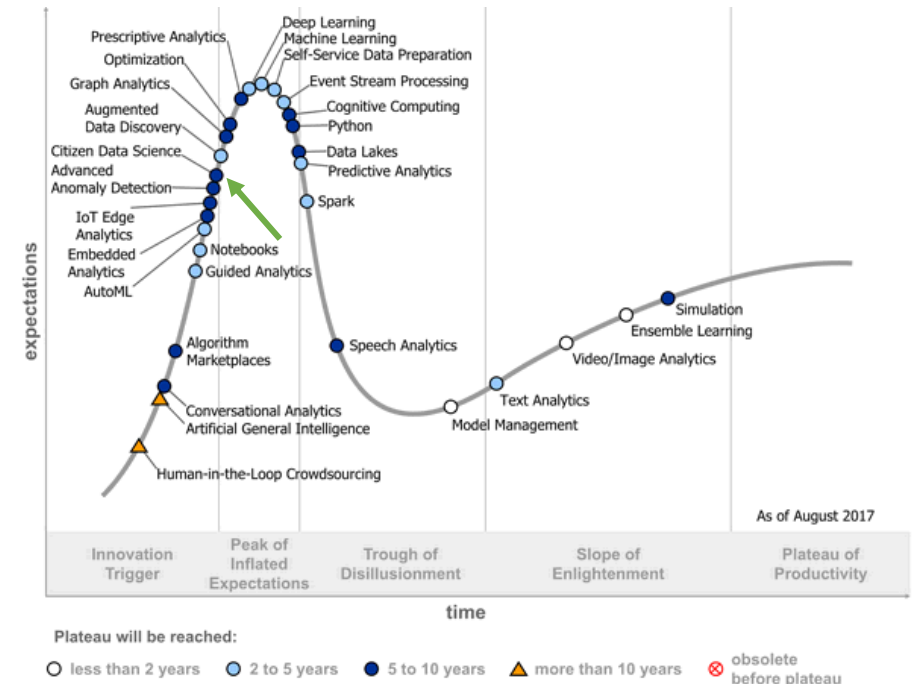
O que é Ciência de Dados?

- Ciência de dados é um termo que **não tem** uma única e completa definição;
- Os métodos, objetivos e aplicações evoluíram com o tempo e avanço da tecnologia;
- 25 atrás: coleta e limpeza de dados e a aplicação de métodos estatísticos;
- Hoje: análise de dados, análise preditiva, mineração de dados, inteligência de negócios, aprendizado de máquina...



O que é Ciência de Dados?

- Os termos “ciência de dados” e “mineração de dados” costumam ser usados alternadamente;
- Ciência de dados ganhou notoriedade à medida que vários indivíduos e organizações tentam tirar proveito do “*hype*” em torno dela;
- A ciência de dados é um **conjunto de princípios fundamentais** que orientam a extração de conhecimento dos dados;
- A mineração de dados é a **extração de conhecimento dos dados**, por meio de tecnologias que incorporam esses princípios;

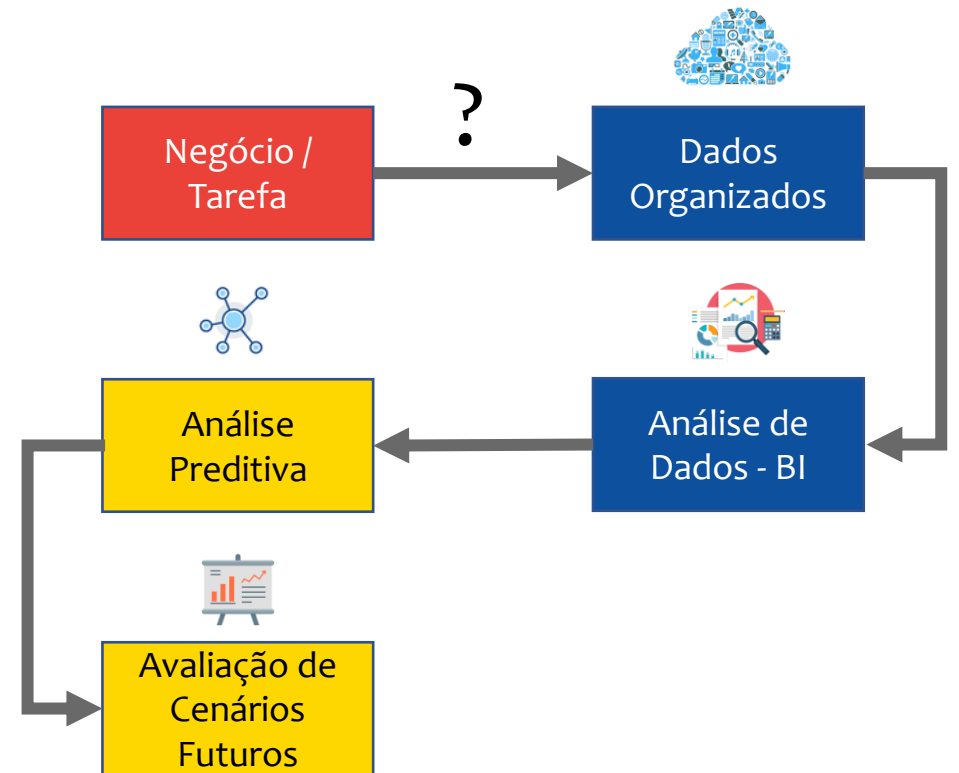


© 2017 Gartner, Inc.

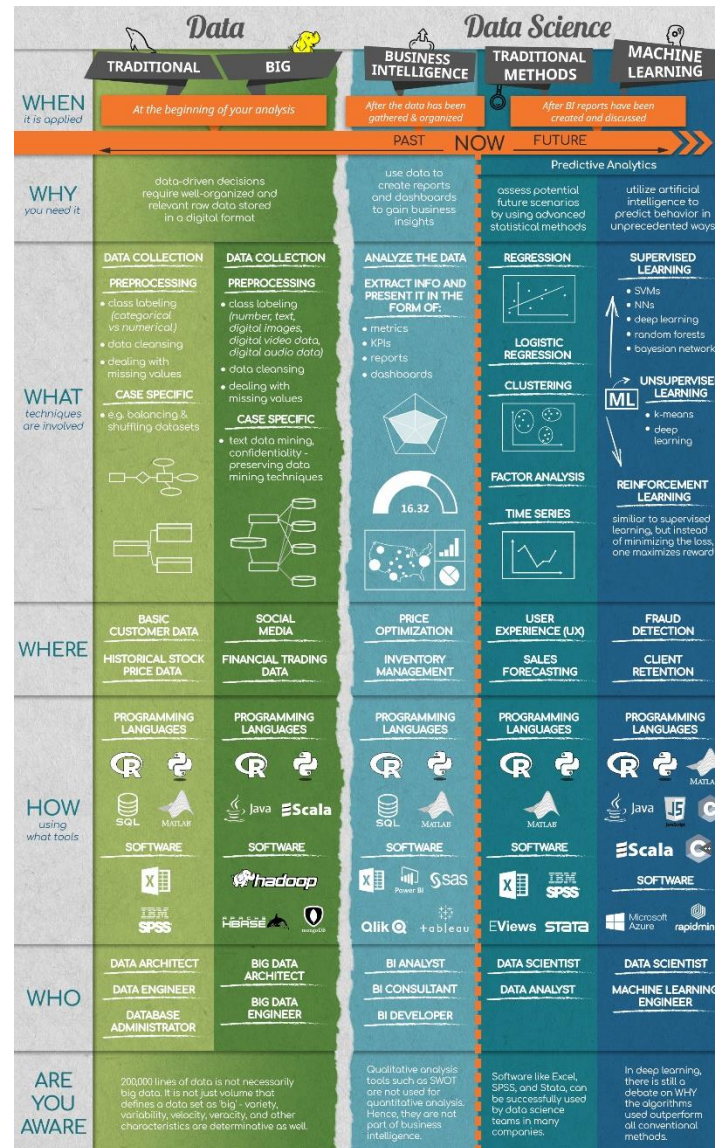
O que é Ciência de Dados?

Como abordar um negócio ou uma tarefa de ciência de dados?

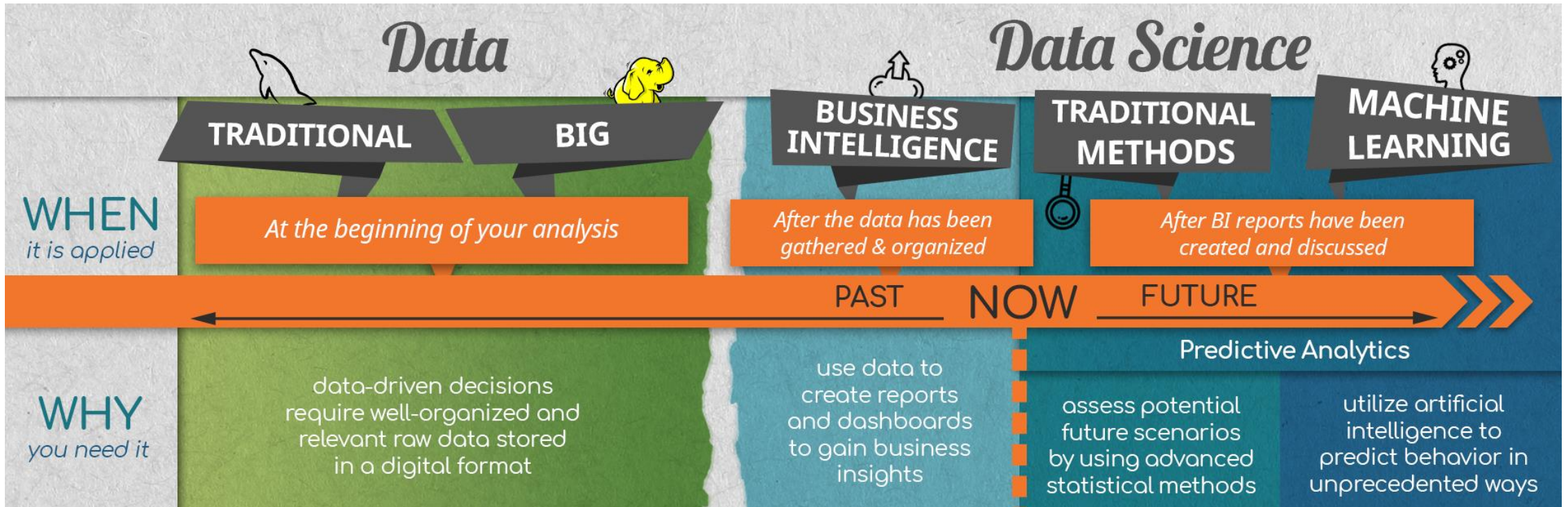
- Cenário 1: você tem os dados adequados e **uma tarefa definida**. Ex.: prever a demanda de um determinado produto para os próximos meses;
- Cenário 2: você recebe uma enorme quantidade de dados e **precisa fazer algo** e **dizer como** aumentar o lucro para o próximo ano;
- Em ambos os casos devemos ter um conjunto de dados adequado;



Definindo Ciência de Dados

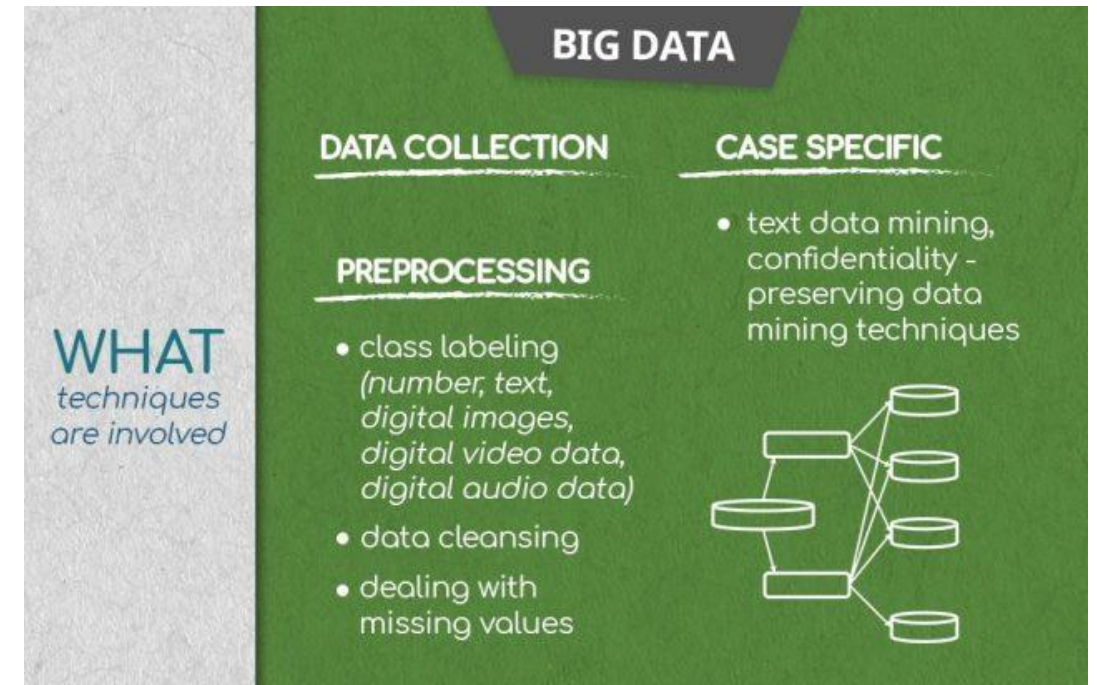
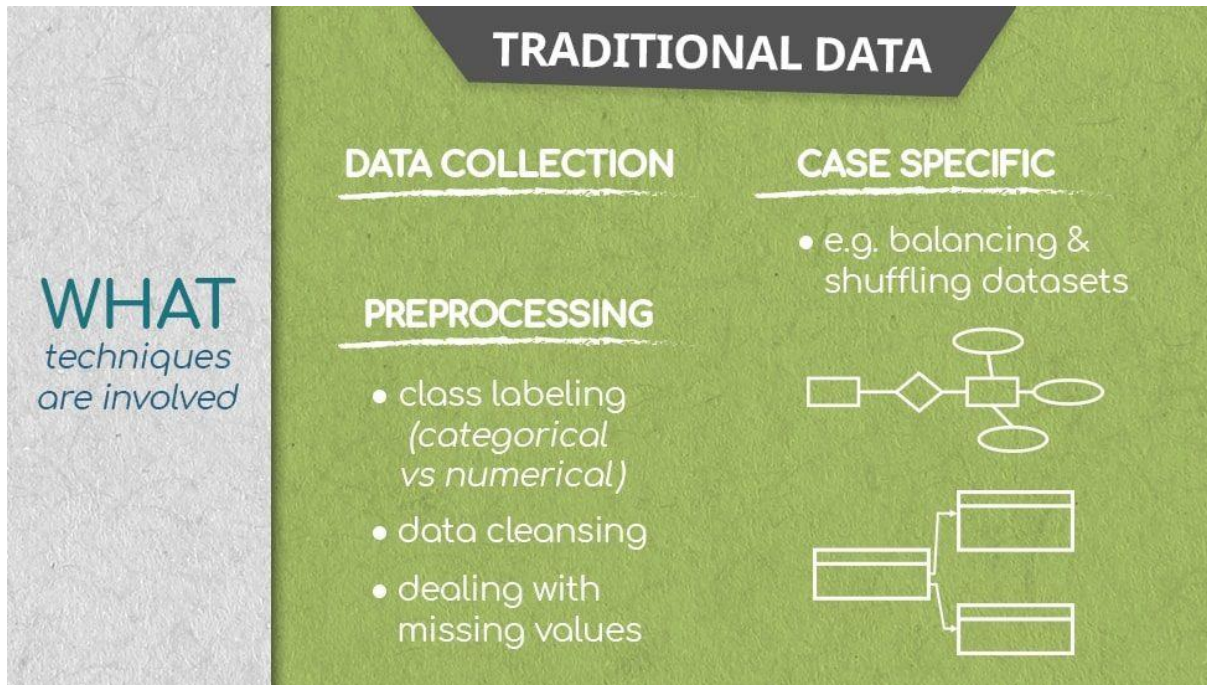


Definindo Ciência de Dados



Os dados

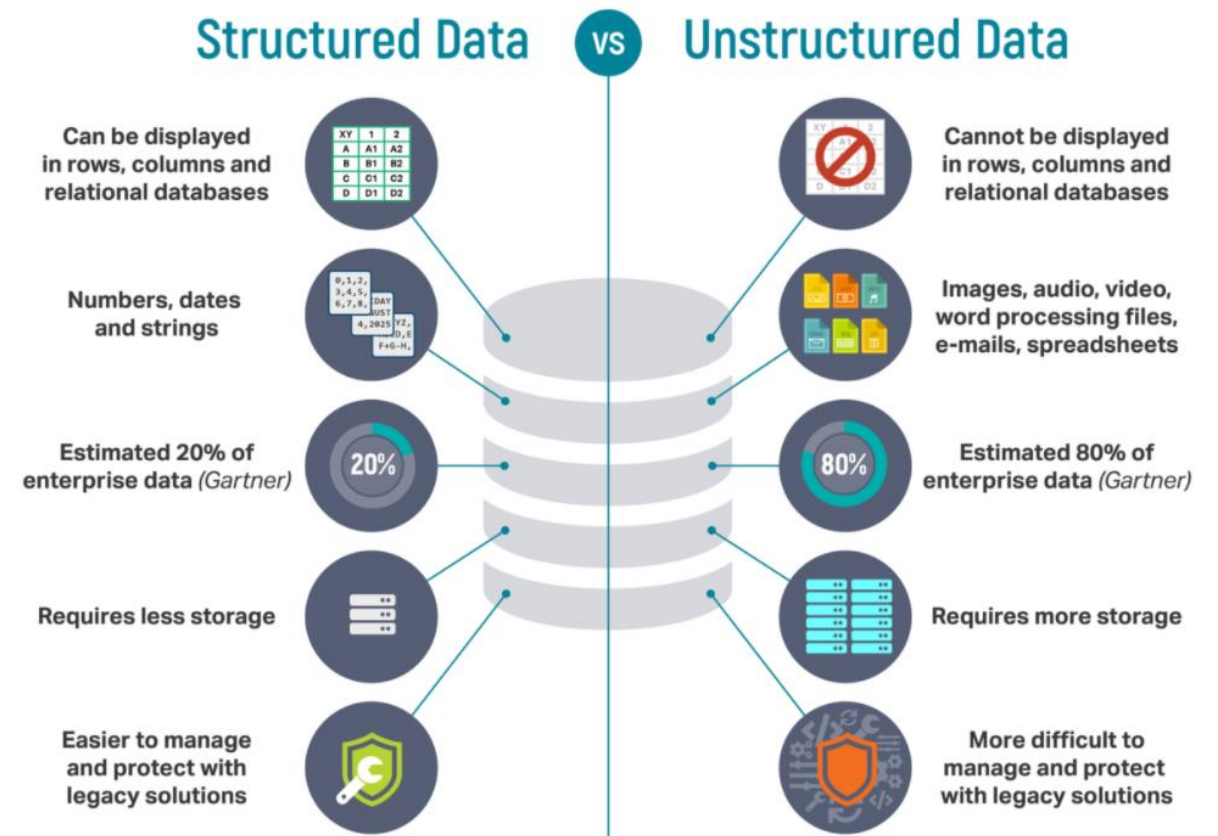
Os dados são a base da ciência de dados



Os dados

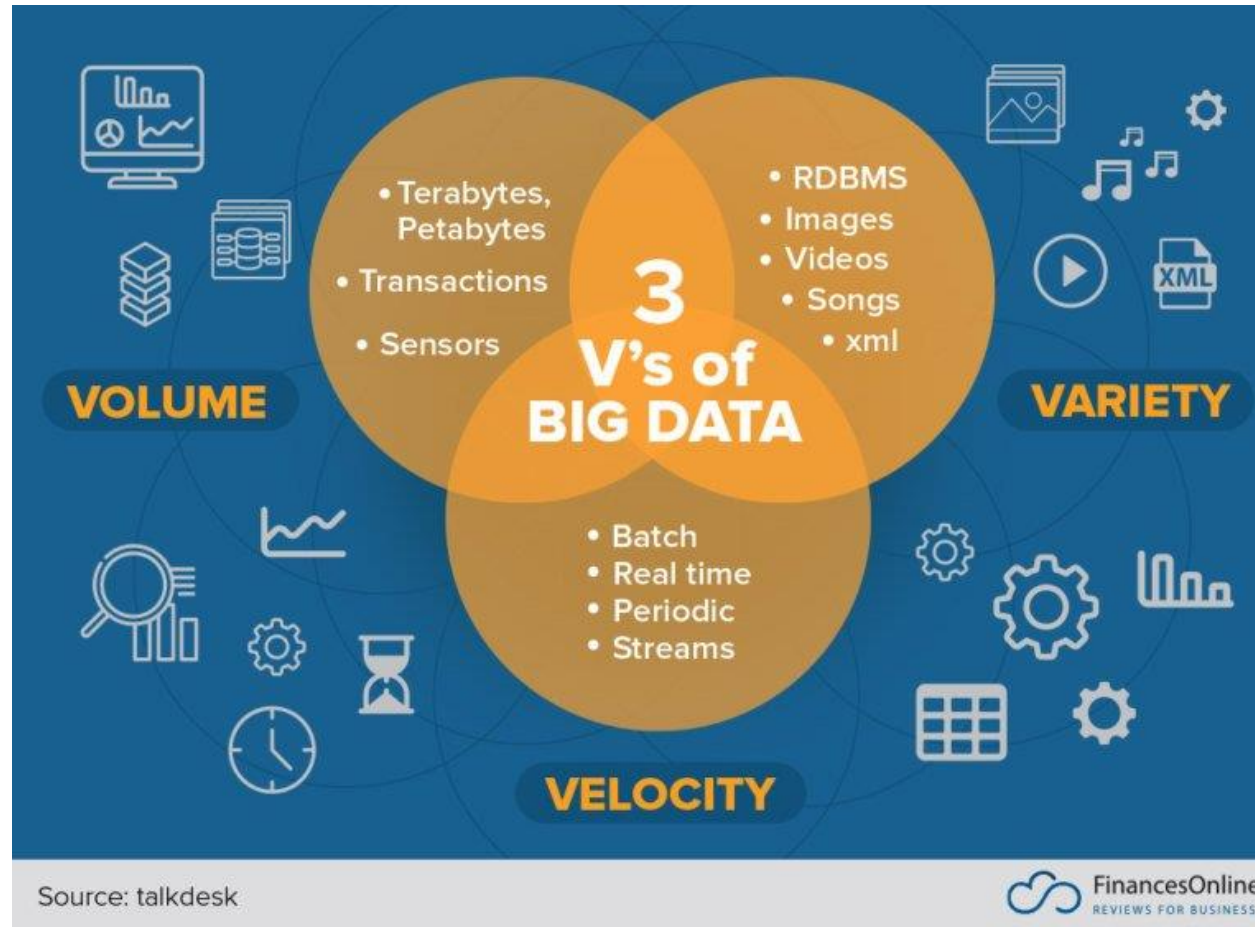
Os dados são a base da ciência de dados

Traditional Data	Big Data
Structured data	Unstructured or Semistructured data
The size of data is very small	The size is more than traditional data size
Data is Centralized	Data are distributed
It is easy to work or manipulate	It is difficult to handle
Normal system configuration is suficiente to process	High system configuration is required to process
A traditional database tools is enough	Special kind os tools are required
Normal functions are enough to manipulate	Requires special kinf of functions to manipulate



Os dados

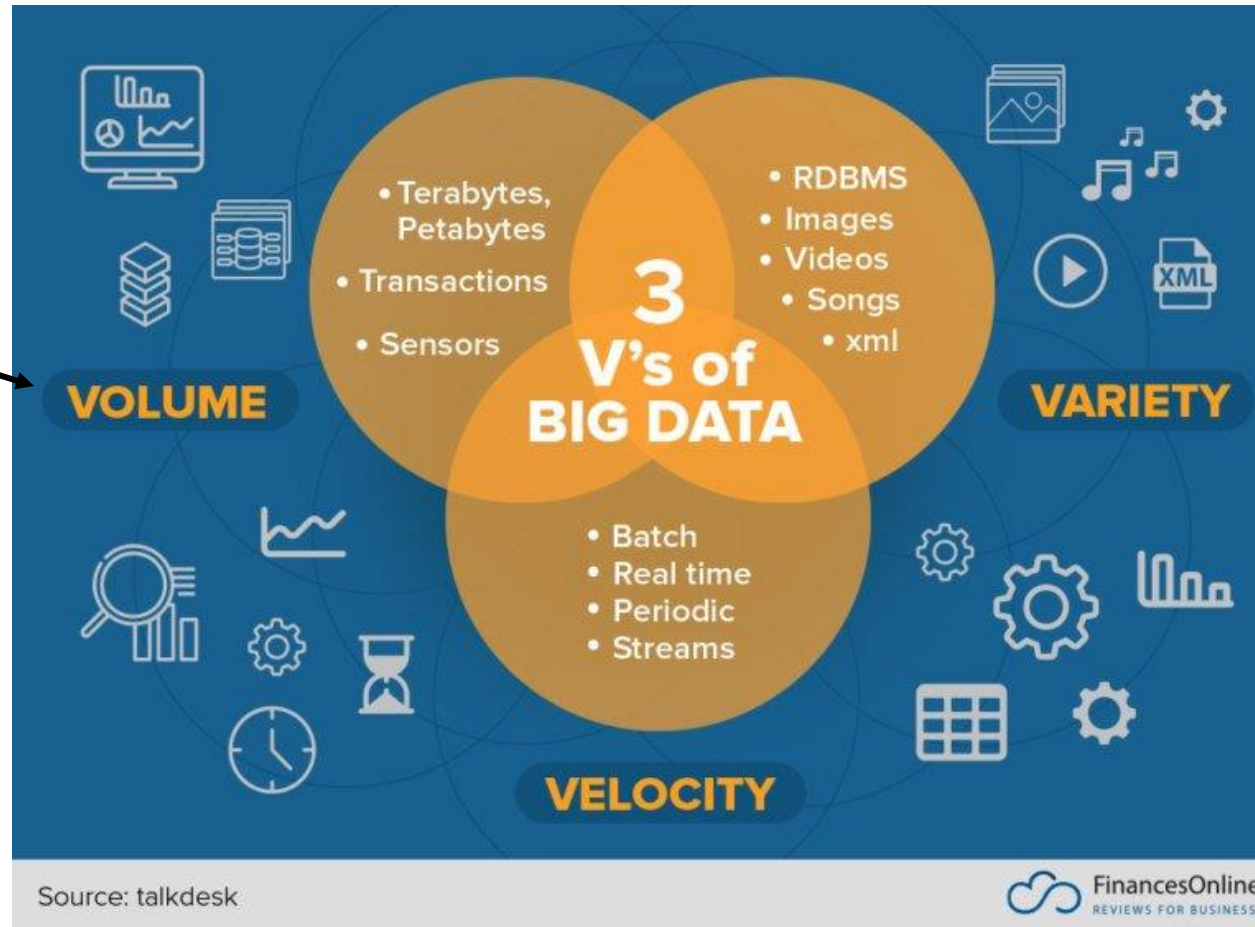
Os dados são a base da ciência de dados



Os dados

Os dados são a base da ciência de dados

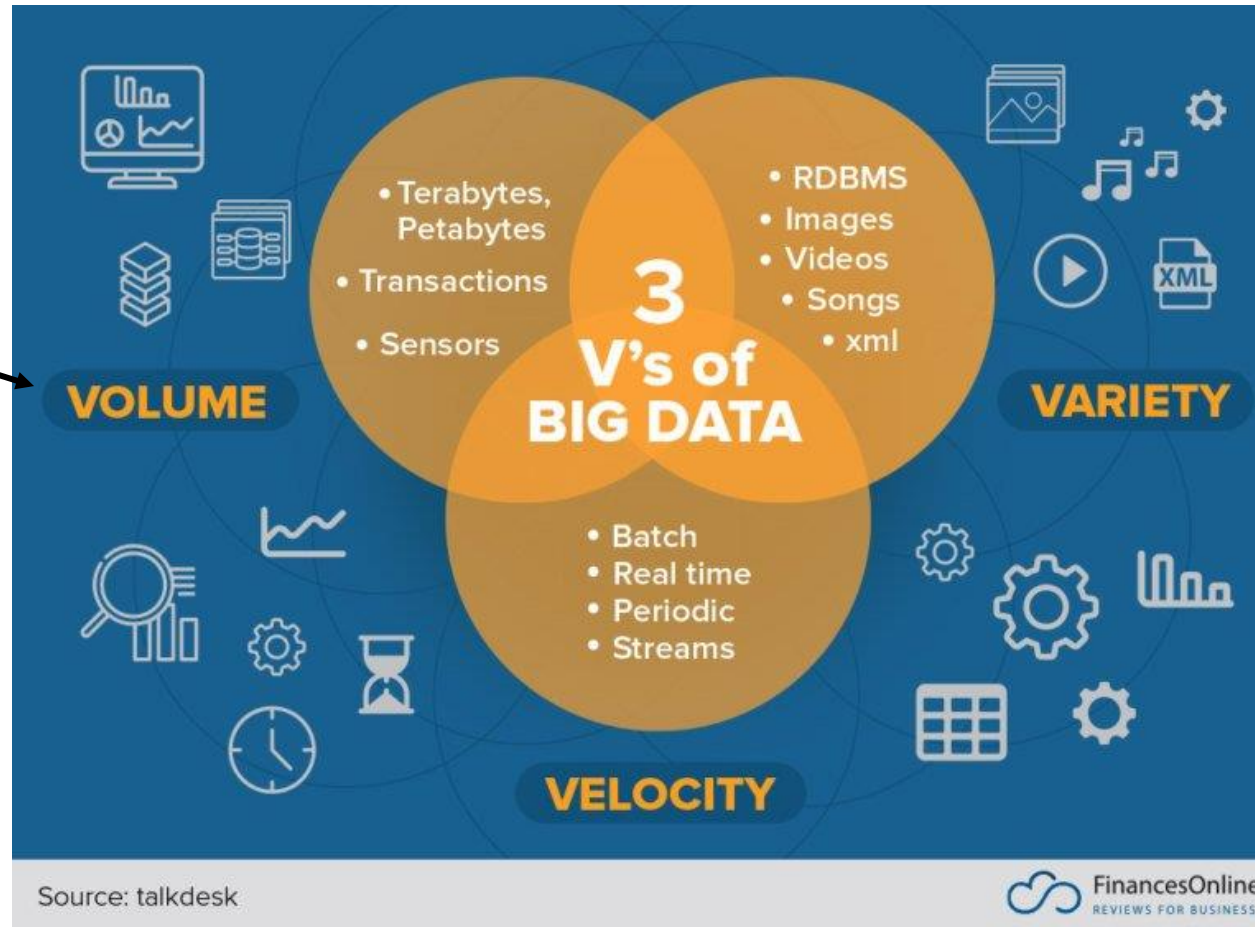
350 milhões de fotos
são enviadas todos os
dias para o Facebook



Os dados

Os dados são a base da ciência de dados

350 milhões de fotos
são enviadas todos os
dias para o Facebook

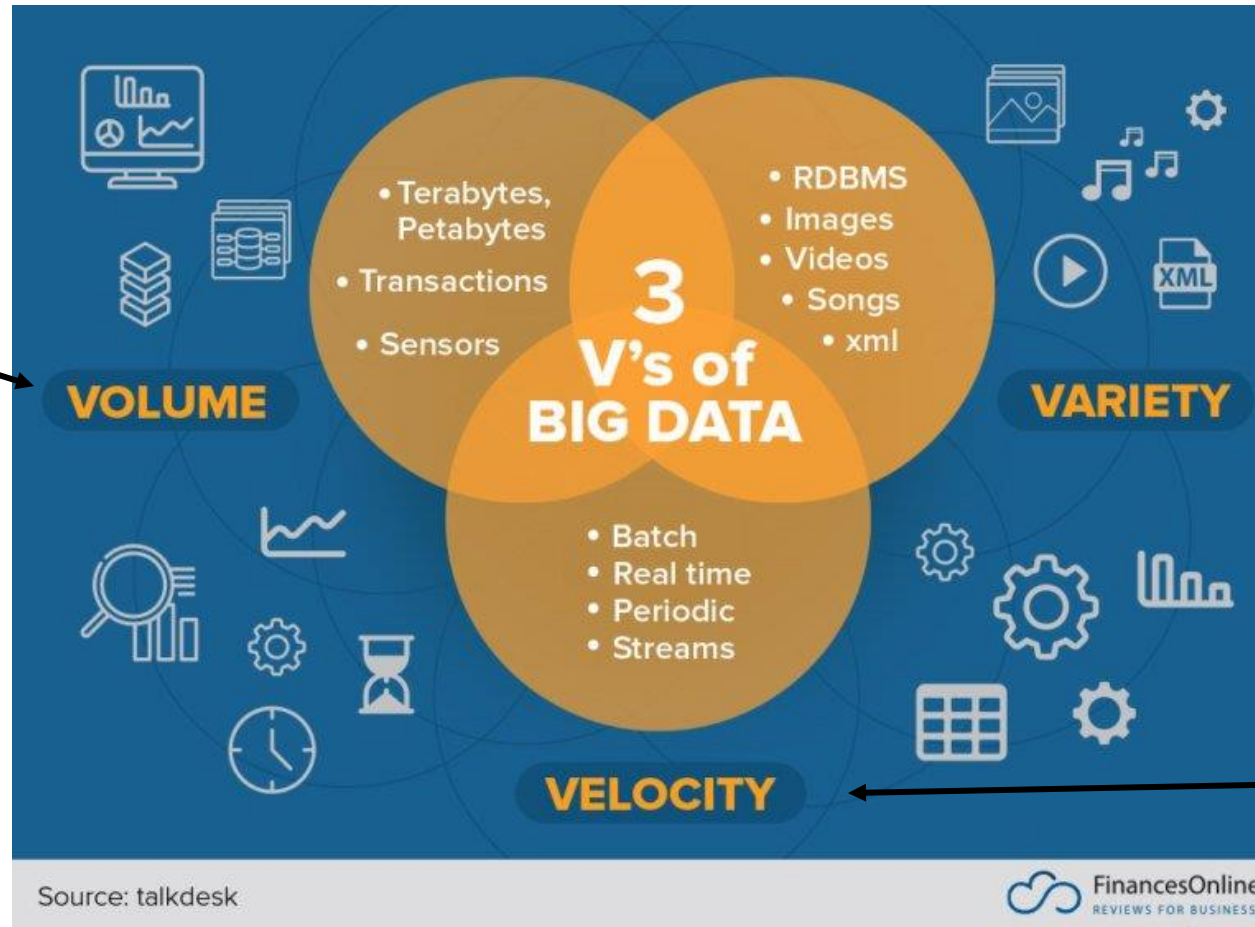


80% deles sendo não
estruturados

Os dados

Os dados são a base da ciência de dados

350 milhões de fotos
são enviadas todos os
dias para o Facebook



80% deles sendo não
estruturados

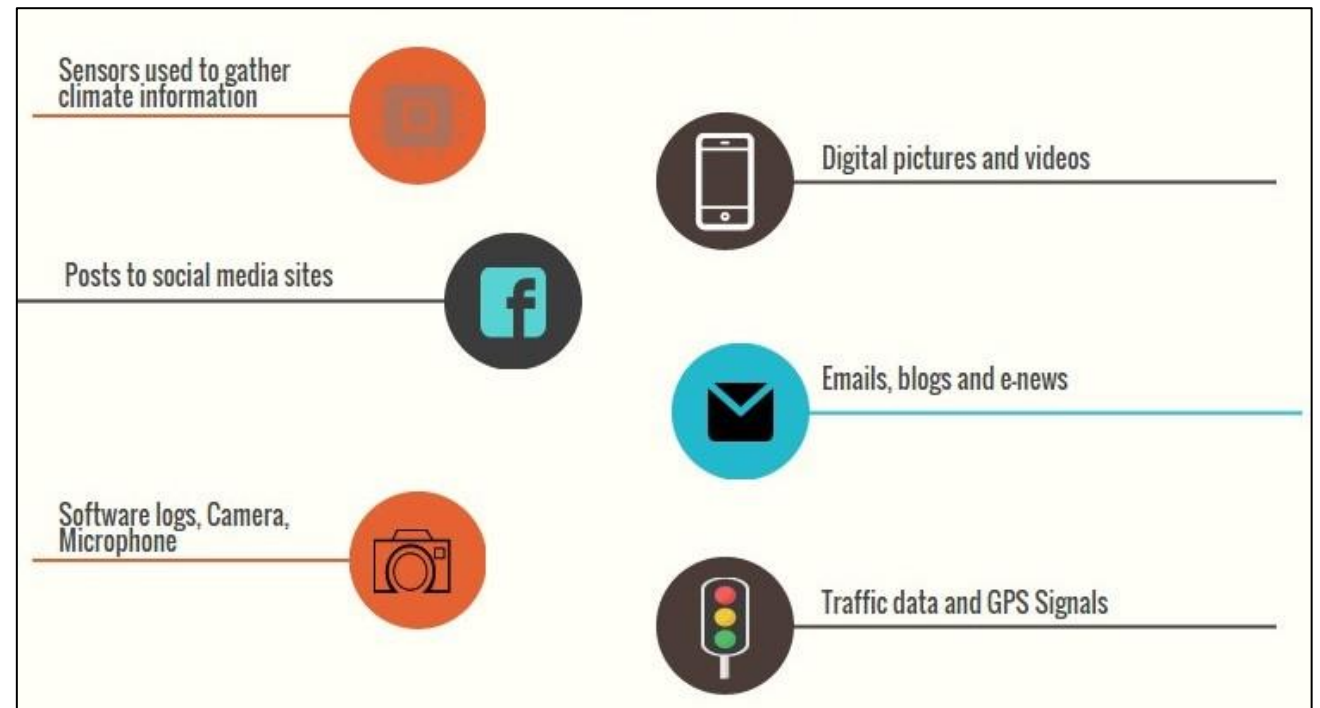
Mais importante de
todos, segundo Jeff
Bezos, CEO da
Amazon

Os dados

De onde vêm os dados?

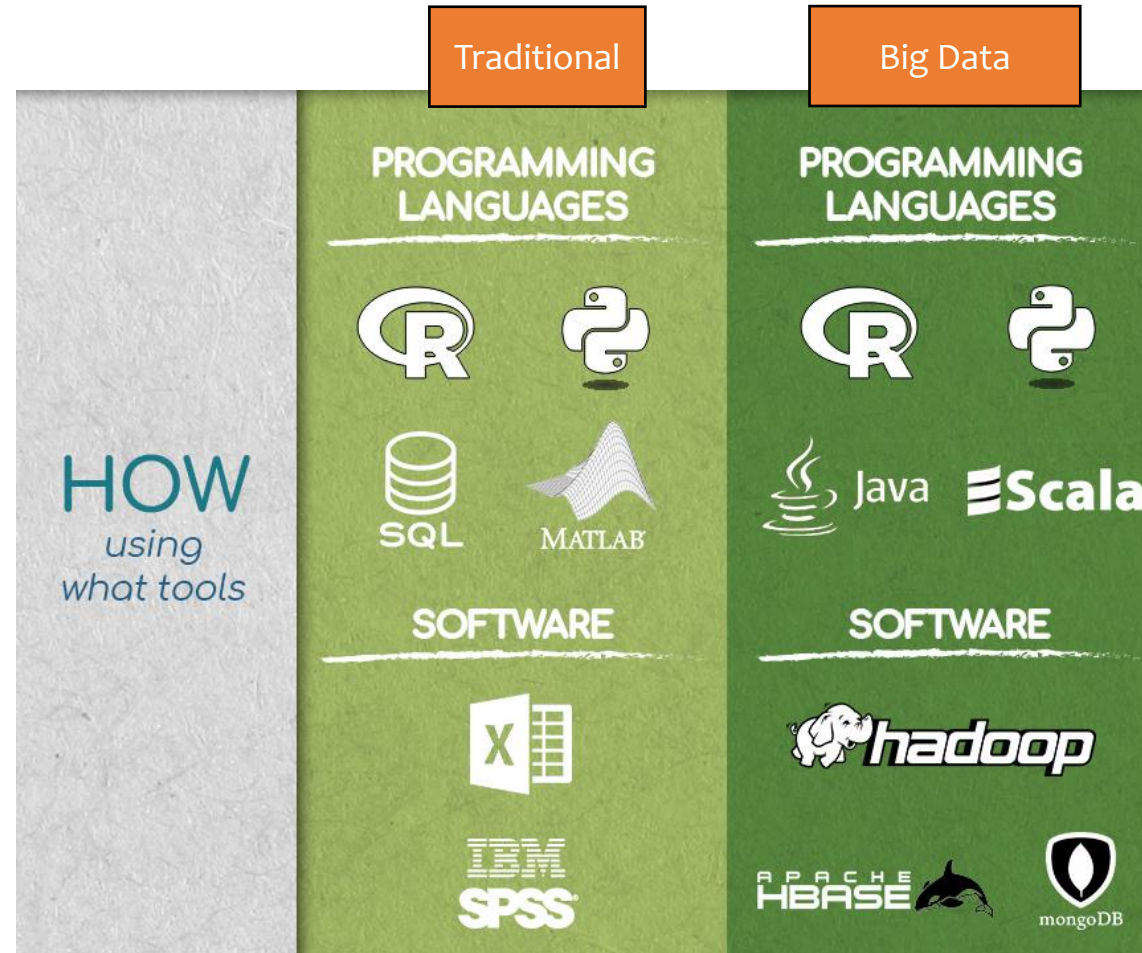


- Facebook: links compartilhados, eventos, pedidos de amizades...
- Twitter: terabytes gerados pelos feeds;
- Dados de GPS de dispositivos móveis;
- Registros de transações e compras;
- IoT (Internet of Things)



Os dados

Como e quais linguagens/softwarewares usar?



Os dados

Quem manipula?



Data architect: projetar, criar, implantar e gerenciar o banco de dados; manter os dados organizados;

Data engineer: realizar o processamento e limpeza dos dados, deixando-os prontos para análise;

Database administrator: controla o fluxo de dados, principalmente com dados **tradicionais**;

Dúvidas?



Bibliografia Básica

- **FAWCETT, T.; PROVOST, F. Data Science para Negócios: O que você precisa saber sobre mineração de dados e pensamento analítico de dados.** Alta Books Editora, 2018.
- **MAYER-SCHÖNBERGER, V.; CUKIER, K. Big data: A revolution that will transform how we live, work, and think.** Houghton Mifflin Harcourt, 2013.
- **GRUS J. Data science from scratch: first principles with python.** O'Reilly Media, 2019.
- **CARVALHO, A. et al. Inteligência Artificial: uma abordagem de aprendizado de máquina.** Rio de Janeiro: LTC. 2011.

Outras Referências

- **Data Science History and Overview.** Disponível em: <https://www.kdnuggets.com/2020/11/data-science-history-overview.html>. Acesso: 15 de março de 2021;
- **Defining Data Science: The What, Where and How of Data Science.** Disponível em: <https://365datascience.com/career-advice/career-guides/defining-data-science/>. Acesso 15 de março de 2021;
- **Data Science: A High-Level Overview.** Disponível em: <https://www.linkedin.com/pulse/data-science-high-level-overview-giuliano-liguori-/>. Acesso: 15 de março de 2021;
- **What is Big Data Analytics and How It Helps You Understand Your Customers. .** Disponível em: <https://financesonline.com/what-is-big-data-analytics-and-how-it-helps-you-understand-your-customers/>. Acesso: 22 de março de 2021