

# Análise de Dados com Base em Processamento Massivo em Paralelo

# Mercado de Engenharia de Dados

André Perez  
ICMC/USP

[andre.marcos.perez@usp.br](mailto:andre.marcos.perez@usp.br)  
[linkedin.com/in/andremarcosperez/](https://www.linkedin.com/in/andremarcosperez/)



# Agenda

- Breve linha do tempo
- *Players* e soluções
- Engenheira(o) de Dados
- Conteúdo extra<sup>2</sup>

# Breve linha do tempo

## ~1970 até 2000

- Poucos *players*
- *Software* de código proprietário
- Escala vertical de soluções
- Infraestrutura *On Premise*

## ~2010 até os dias de hoje

- Muitos *players*
- *Software* de código aberto
- Escala horizontal das soluções
- Infraestrutura *On Premise* e *Cloud*



## ~2000 até 2010

- Poucos *players*
- *Software* de código aberto
- Escala horizontal das soluções
- Infraestrutura *On Premise*



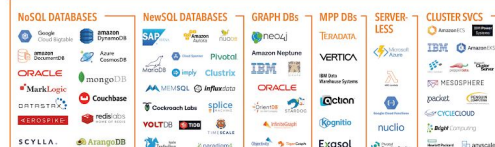
## INFRASTRUCTURE



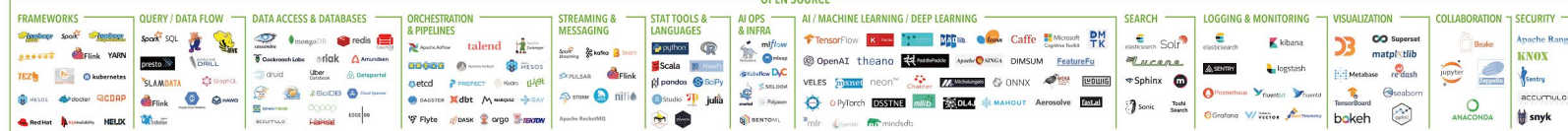
## ANALYTICS &amp; MACHINE INTELLIGENCE



## APPLICATIONS - ENTERPRISE



## OPEN SOURCE



## DATA SOURCES &amp; APIs



## DATA RESOURCES



Fonte: <https://mattturck.com/data2020/>

2020

# OPEN SOURCE

- Serviços gerenciados;
- Consultoria;
- Suporte 24/7;
- Treinamento;
- Certificação;
- etc.

# OPEN SOURCE

- Serviços gerenciados;
- Consultoria;
- Suporte 24/7;
- Treinamento;
- Certificação;
- etc.



## INFRASTRUCTURE

**STORAGE**

**HADOOP**

**CLOUDSERA**

**DATA LAKES**

**DATA WAREHOUSES**

**STREAMING / IN-MEMORY**

## ANALYTICS &amp; MACHINE INTELLIGENCE

**BI PLATFORMS**

**VISUALIZATION**

**DATA ANALYST PLATFORMS**

## APPLICATIONS - ENTERPRISE

**SALES**

**MARKETING - B2B**

**MARKETING - B2C**

**CUSTOMER EXPERIENCE / SERVICE**

**HUMAN CAPITAL**

**NoSQL DATABASES**

**NewSQL DATABASES**

**GRAPH DBs**

**MPP DBs**

**SERVER-LESS**

**CLUSTER SVCS**

**DATA SCIENCE NOTEBOOKS**

**DATA SCIENCE PLATFORMS**

**MACHINE LEARNING**

**ETL / DATA TRANSFORMATION**

**DATA INTEGRATION**

**DATA GOVERNANCE**

**DATA QUALITY**

**COMPUTER VISION**

**HORIZONTAL AI**

**SPEECH & NLP**

**MGMT / MONITORING**

**DATA GENERATION & LABELLING**

**AI OPS**

**GPU DBs & CLOUD**

**AI HARDWARE**

**SEARCH**

**LOG ANALYTICS**

**SOCIAL ANALYTICS**

**WEB / MOBILE / COMM-ANALYTICS**

**FRAMEWORKS**

**QUERY / DATA FLOW**

**DATA ACCESS & DATABASES**

**ORCHESTRATION & PIPELINES**

**STREAMING & MESSAGING**

**STAT TOOLS & LANGUAGES**

**AI OPS & INFRA**

**AI / MACHINE LEARNING / DEEP LEARNING**

## OPEN SOURCE

**SEARCH**

**LOGGING & MONITORING**

**VISUALIZATION**

**COLLABORATION**

**SECURITY**

**DATA MARKETPLACES & DISCOVERY**

**FINANCIAL & ECONOMIC DATA**

**AIR / SPACE / SEA**

**PEOPLE / ENTITIES**

**LOCATION INTELLIGENCE**

**OTHER**

**DATA SOURCES & APIs**

**DATA RESOURCES**

**ADVERTISING**

**EDUCATION**

**REAL ESTATE**

**COMMERCE**

**FINANCE - LENDING**

**INSURANCE**

**HEALTHCARE**

**LIFE SCIENCES**

**TRANSPORTATION**

**AGRICULTURE**

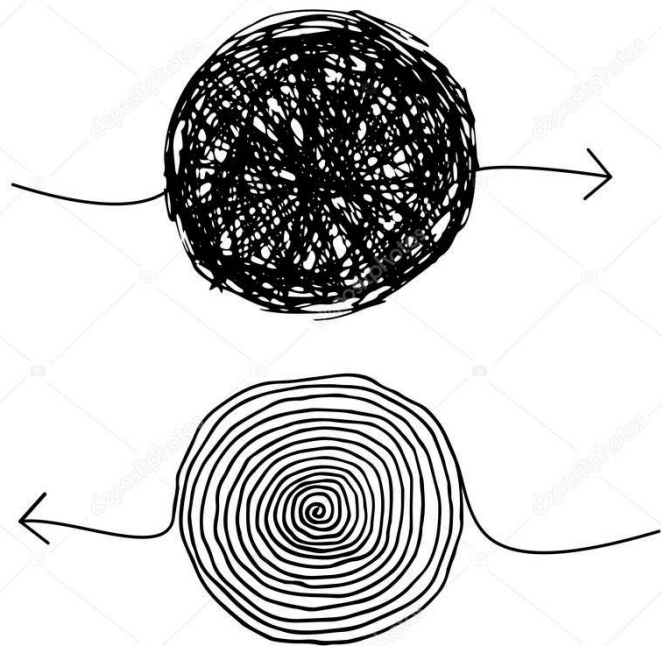
**INDUSTRIAL**

**OTHER**

Fonte: <https://mattturck.com/data2020/>

2020

# Engenheira(o) de Dados





# Engenheira(o) de Dados x Cientista de Dados

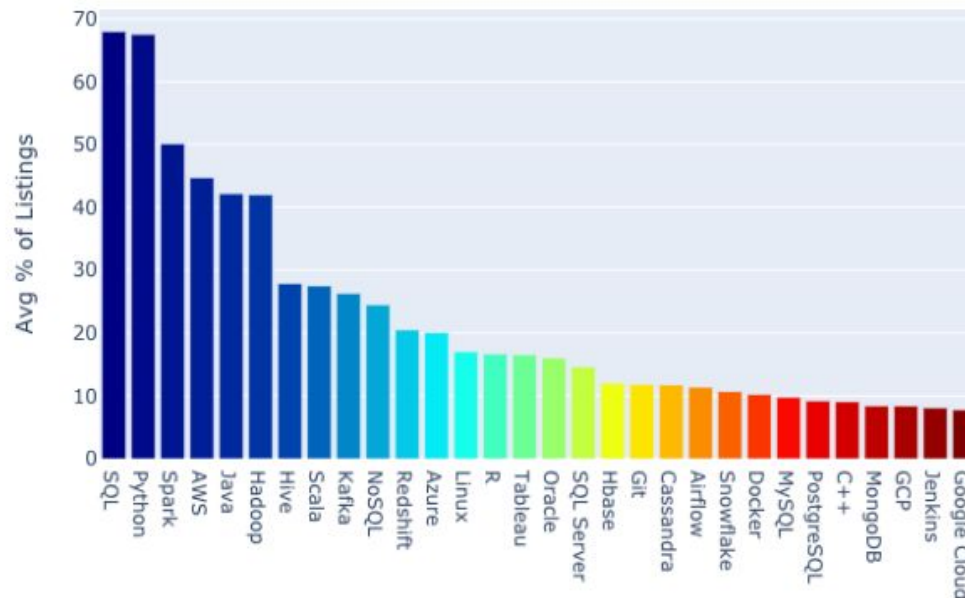
2 a 5

Estima-se que são necessários de 2 a 5 engenheiros de dados para cada cientista de dados para que um time de dados funcione bem.

Fonte: <https://www.oreilly.com/radar/data-engineers-vs-data-scientists/>

# Engenheira(o) de Dados x Cientista de Dados

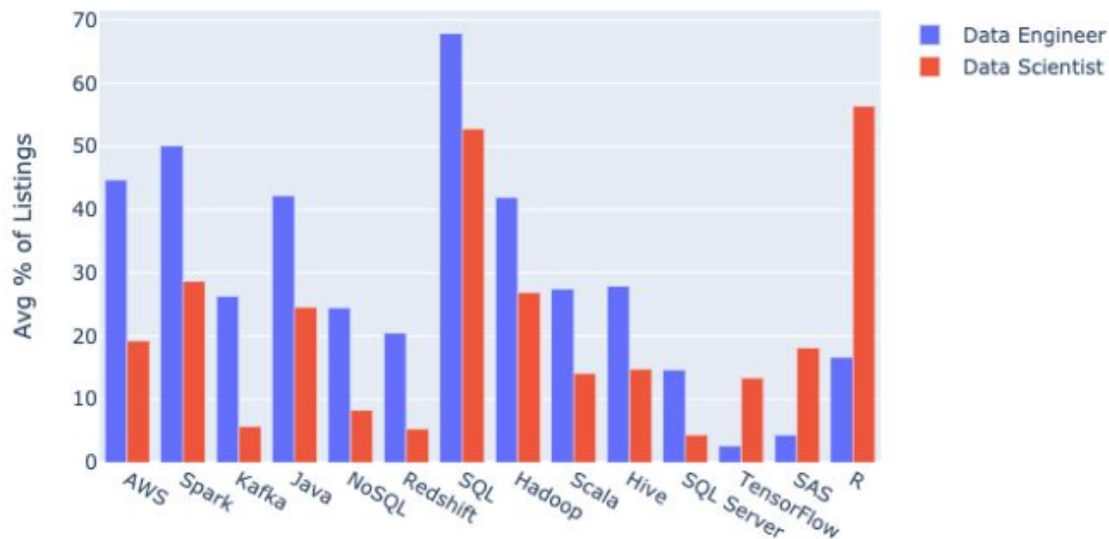
Technologies in Data Engineer Job Listings 2020



Fonte: <https://towardsdatascience.com/most-in-demand-tech-skills-for-data-engineers-58f4c1ca25ab>

# Engenheira(o) de Dados x Cientista de Dados

Largest Differences for Data Engineer vs. Data Scientist Job Listing Keywords



Fonte: <https://towardsdatascience.com/most-in-demand-tech-skills-for-data-engineers-58f4c1ca25ab>



# Conteúdo extra<sup>2</sup>

## Artigos

- [The Rise of the Data Engineer](#) (Medium - 20/01/17)
- [A beginner's Guide to Data Engineering](#) (Medium - 08/01/18)

## Podcasts

- [O que é Data Lake](#) (Data Hackers - 04/09/20)
- [Engenharia de dados na Serasa Experian](#) (Hipsters.Tech - 17/07/20)
- [Engenharia de Dados no Grupo Zap](#) (Data Hackers - 20/09/19)
- [Tecnologias no Nubank: 3 anos depois](#) (Hipsters.Tech - 28/05/19)
- [Nubank Data Engineering with Sujith Nair](#) (SED - 06/02/20)