



CIÊNCIA DE DADOS: VISÃO GERAL E FORMAÇÃO



Eduardo Ogasawara
eogasawara@ieee.org
<https://eic.cefet-rj.br/~eogasawara>

Biografia

- Doutorado em Sistemas e Computação (COPPE/UFRJ) em 2011
- Professor da Escola de Informática e Computação (EIC) - CEFET/RJ
 - Departamento de Ciência da Computação (DEPIN)
 - Coordenação de Informática (COINFO)
- Docente permanente
 - Programa de Pós-graduação em Ciência da Computação (PPCIC)
 - Programa de Pós-graduação em Engenharia de Produção e Sistemas (PPPRO)
- Membro da IEEE, SBC, ACM e INNS
- Representante Institucional da SBC



eogasawara@ieee.org
<https://eic.cefet-rj.br/~eogasawara>

Informes

Simpósio Brasileiro de Banco de Dados (SBBD)



Inscreeva-se!

SBBD 2021 – BRAZILIAN SYMPOSIUM ON DATABASES

The Premier Brazilian Conference on Data Science and Big Data

The annual **Brazilian Symposium on Databases (SBBD)** is the official event on databases of the Brazilian Computer Society (SBC). The symposium includes a technical program with research and industrial talks, tutorials, demos, and focused workshops. It also hosts invited talks by distinguished speakers from the international research community.

Due to COVID-19 and coronavirus pandemic, all activities of the 36th edition of the SBBD will happen **online** only, **October 4-8, 2021** – organized by [CEFET/RJ](#) (Rio de Janeiro, Brazil).

REALIZATION	EXECUTION	ACADEMIC SUPPORT
		
SPONSOR	SUPPORT	
 		

TRACKS COORDINATION

Program Chair: Ricardo Torres (NTNU, Norway)
Short Vision Industrial Chair: Damires Souza (IFPB, Brazil)
Steering Committee Chair: Fabio Porto (LNCC, Brazil)
Demos and Applications Chair: Leonardo Ribeiro (UFG, Brazil)
Thesis and Dissertation Workshop Chair (WTDBD): Julio Reis (Unicamp, Brazil)
CTDBD Chair: Cristina Ciferri (USP, Brazil)
Short courses Chair: Alessandréia M de Oliveira (UFJF, Brazil)
Tutorials Chair: Daniel de Oliveira (UFF, Brazil)
Workshop Chair: Eduardo Almeida (UFPR, Brazil)
WTAG Chair: André Carvalho (UFAM, Brazil)

ONLINE ORGANIZATION

SBBD General Chair:
Eduardo Ogawasara (CEFET/RJ) – eduardo.ogawasara@cefet-rj.br
Rafaeli Coutinho (CEFET/RJ) – rafaelli.coutinho@cefet-rj.br

Painel de Formação em Ciência de Dados (All we need is Data)



Série de lives promovidas pelo CD-CEBD



APRESENTAM

O CIENTISTA DE DADOS E OUTROS UNICÓRNIOS



Karin Becker
UFRGS
mediadora



Ângelo Brayner
UFC
convidado

Rodrigo Barros
PUCRS/Teia Labs
convidado



Live no Youtube

21/09/2021 | 17h (Brasil)

youtube.com/SociedadeBrasileiradeComputação

Vamos começar...



2012

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

BUSINESS INSIDER CAREERS

Here's how much money you make in the 'sexiest job of the 21st century'

Jacquelyn Smith     

Feb. 25, 2016, 3:20 PM  33,395

In 2012, Harvard Business Review named data scientist the "[sexiest job of the 21st century](#)." More recently, Glassdoor named it the "[best job of the year](#)" for 2016.

"It isn't a big surprise," Dr. Andrew Chamberlain, Glassdoor's chief economist, told Business Insider. "It's one of the hottest and fastest growing jobs we're seeing right now."

According to Glassdoor, data scientists earn a base pay of **\$116,840 a year**, on average.



VFS Digital Design/Flickr

Facebook: \$133,841

Apple: \$149,963

Airbnb: \$117,229

Twitter: \$134,861

Microsoft: \$119,129

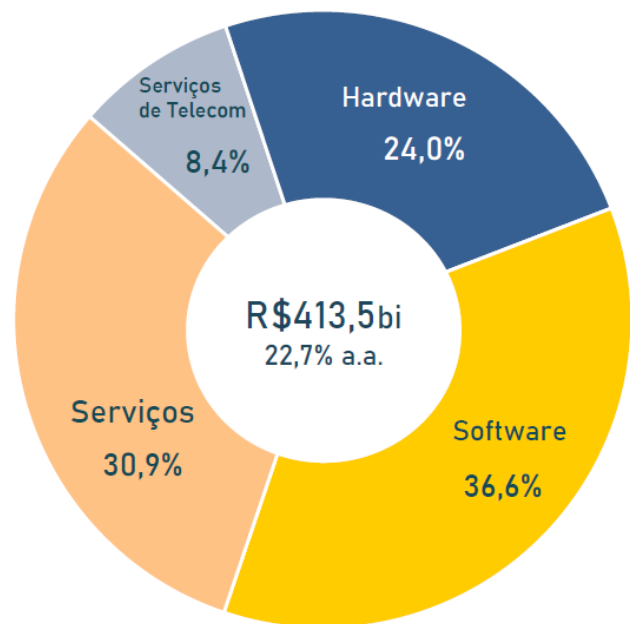
LinkedIn: \$138,798

Perspectivas de Investimentos em TI

Perspectivas de Investimentos de 2021-2024 (R\$ bilhões)



Tecnologias de Transformação Digital



Nuvem
R\$ 181,9 bi | 28% a.a.



Robótica
R\$ 31,4 bi | 6% a.a.



Internet das Coisas
R\$ 74,3 bi | 29% a.a.



Realidade Virtual
R\$ 2,2 bi | 40% a.a.



Big Data & Analytics
R\$ 77,2 bi | 13% a.a.



Redes Sociais
R\$ 14,1 bi | 19% a.a.



Segurança da Informação
R\$ 16,2 bi | 13% a.a.



Impressão 3D
R\$ 0,7 bi | 15% a.a.



Inteligência Artificial
R\$ 11,5 bi | 31% a.a.



Blockchain
R\$ 3,8 bi | 65% a.a.

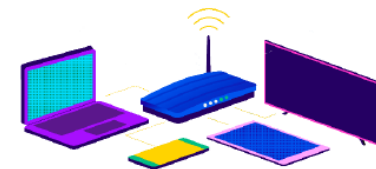
Mobilidade e Conectividade

Mobile, Dados e Banda Larga



R\$ 431,5 bi

4,6% a.a.



Taxa de câmbio: R\$/US\$ 3,95 (2019)

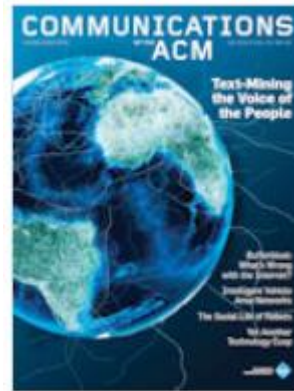
Contexto de Ciência de Dados



ago/2002



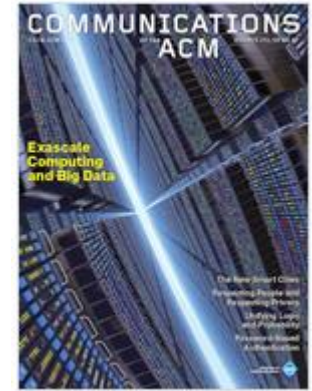
dez/2008



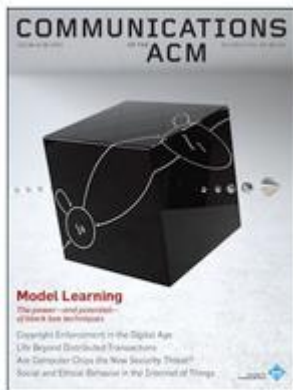
fev/2012



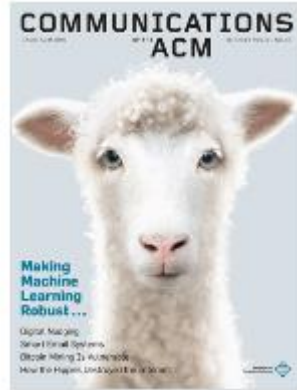
dez/2013



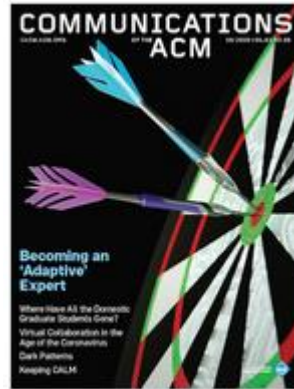
jul/2015



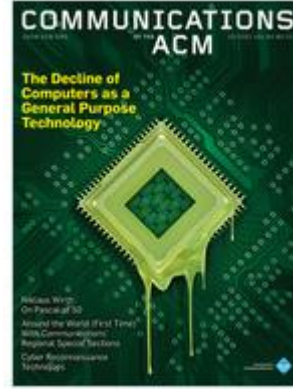
fev/2017



jul/2018



set/2020



mar/2021



jul/2021

Ciência de Dados

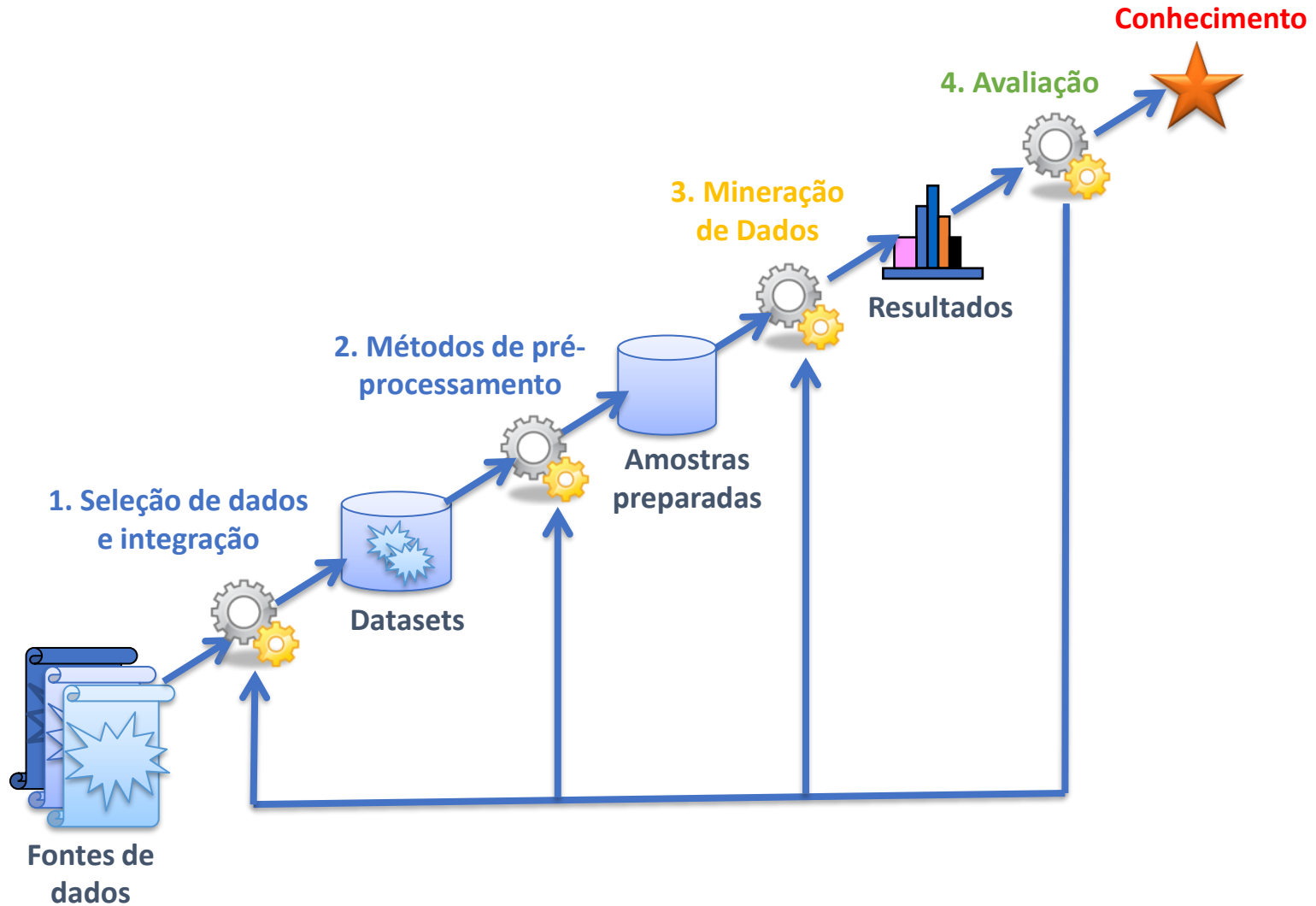
- Ciência de dados é o estudo da extração de conhecimento a partir de dados (Big Data)
- Envolve
 - Análise de dados
 - Aprendizagem estatística e de máquina
 - Gerência de grandes coleções de dados
 - Mineração de dados
 - Computação paralela e distribuída
 - Visualização de dados
- Característica multidisciplinar
- Aplicações na ciência/indústria/governo
 - Bioinformática, petróleo, energia, finanças, astronomia, Internet, mobilidade urbana, defesa cibernética, educação, etc

[1] V. Dhar, 2013, Data science and prediction, *Communications of the ACM*, v. 56, n. 12, p. 64–73.

[2] L. Cao, 2017, Data science: A comprehensive overview, *ACM Computing Surveys*, v. 50, n. 3

[3] L. Cao, 2017, Data science: Challenges and directions, *Communications of the ACM*, v. 60, n. 8, p. 59–68.

Processo de Ciência de Dados

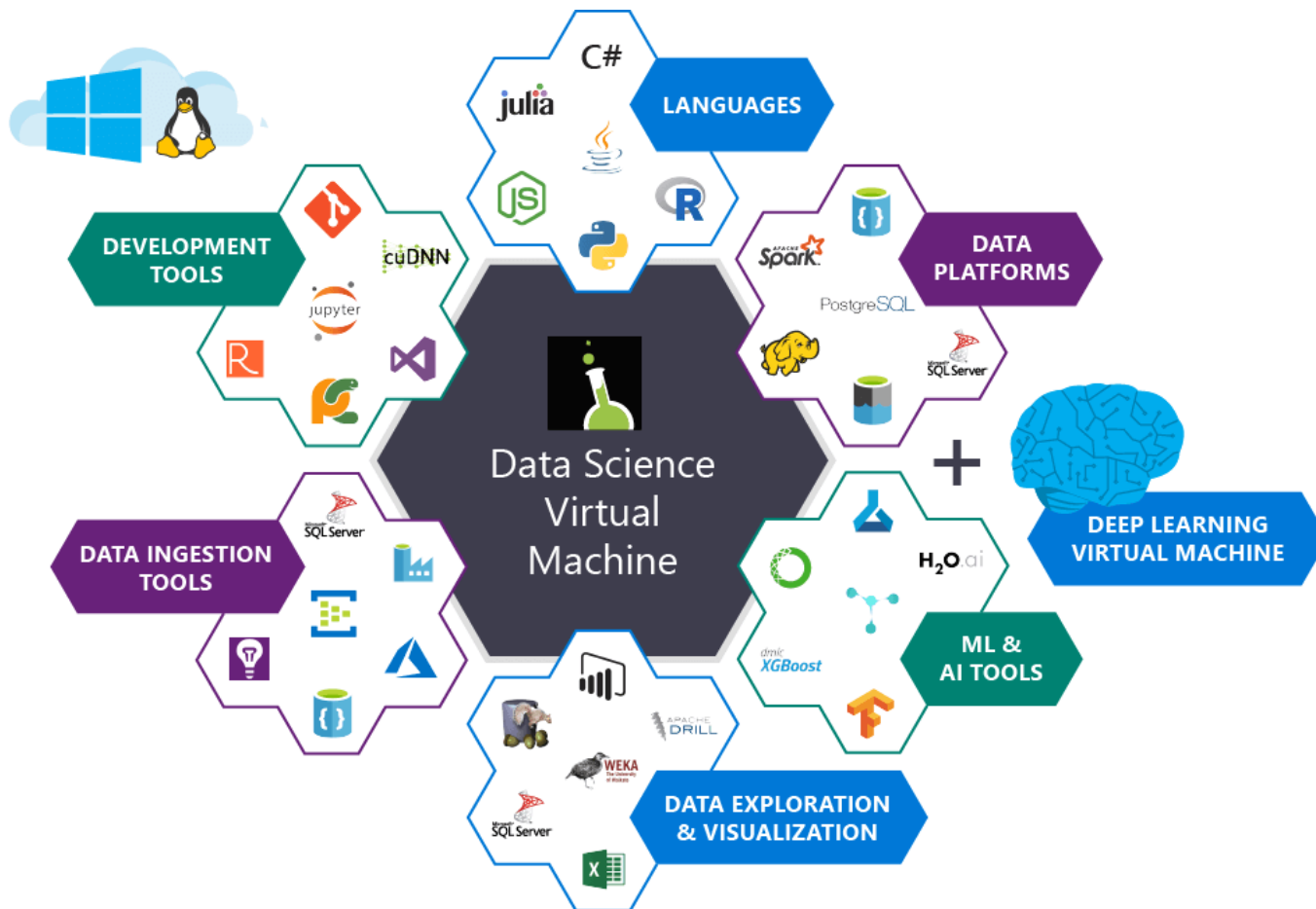


[1] V. Stodden, 2020, The data science life cycle, *Communications of the ACM*, v. 63, n. 7, p. 58–66.

[2] K. Matsudaira, 2015, The science of managing data science, *Communications of the ACM*, v. 58, n. 6, p. 44–47.

[3] J. Han, J. Pei, and M. Kamber, 2011, *Data Mining: Concepts and Techniques*. Elsevier.

Ciência de Dados (ferramentas)



O que se espera do Cientista de Dados?

- Teoria e Prática
 - Fundamentos de matemática e estatística
 - Banco de dados
 - Programação paralela e distribuída
 - Mineração de dados
 - Visualização
 - Conhecimento do domínio
 - Inovação
- Desenvolvimento de software
 - Diversas linguagens de programação
 - Diversas ferramentas

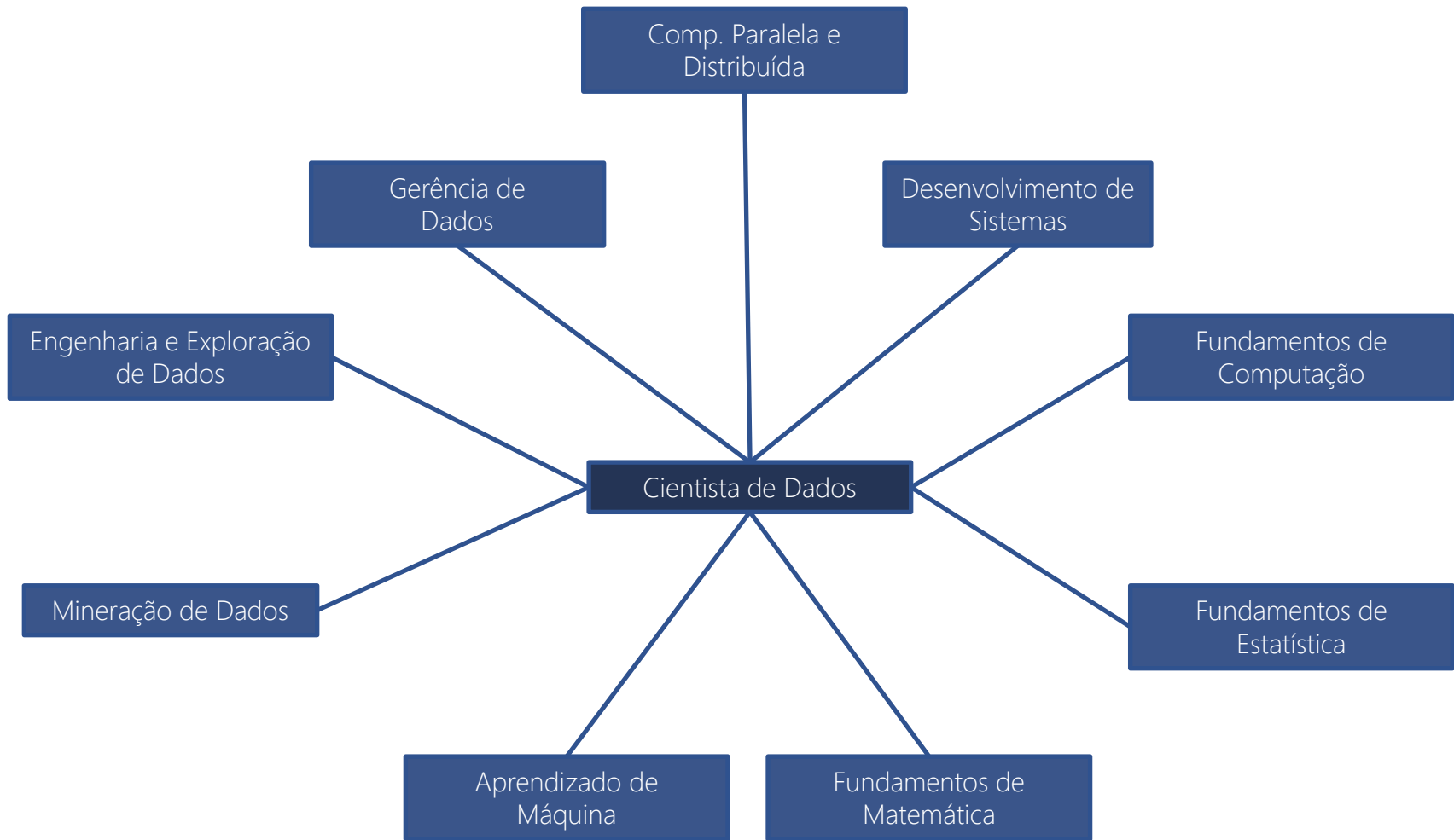


[1] H.B. Barua and K.C. Mondal, 2019, A comprehensive survey on cloud data mining (CDM) frameworks and algorithms, *ACM Computing Surveys*, v. 52, n. 5

[2] D.A. Reed and J. Dongarra, 2015, Exascale computing and big data, *Communications of the ACM*, v. 58, n. 7, p. 56–68.

[3] V. Stodden, 2020, The data science life cycle, *Communications of the ACM*, v. 63, n. 7, p. 58–66.

Graduação em Ciência da Computação



Estatística versus Computação?

[1] ACM Data Science Task Force, *ACM Issues Undergraduate Data Science Curricula*, <https://www.acm.org/media-center/2021/march/data-science-curricula-2021>.

[2] P.J. Diggle, 2015, Statistics: A data science for the 21st century, *Journal of the Royal Statistical Society. Series A: Statistics in Society*, v. 178, n. 4, p. 793–813.

[3] Comissão da SBC para graduação em Ciência de Dados (CEBD)

Programas de Pós-graduação em Ciência de Dados

- Estudo realizado em 2015
 - Predominância nos EUA
 - Foram estudados mais de vinte programas
 - Organização e disciplinas oferecidas
- Área Interdisciplinar
 - Interação com diversos departamentos
- Sólida ligação com
 - Ciência da Computação
 - Estatística
 - Área de Aplicação (interface)

- Mestrado acadêmico criado desde 2016
- Área de Ciência da Computação
- Perfil do Egresso:
 - Sólida formação em Ciência da Computação
- Diferencial do Programa:
 - Formação de Cientistas de Dados
- Egresso deve estar apto a:
 - Atuar em setores onde há necessidade de extração de conhecimento a partir de dados
 - Tratar de questões teórico-práticas relacionadas aos problemas mais atuais relacionados à Computação
 - Propagar conhecimento no meio acadêmico-científico

Disciplinas

Básicas

- Análise e Projeto de Algoritmos
- Arquitetura de Computadores
- Banco de Dados
- Computação Paralela e Distribuída
- Metodologia Científica em Computação
- Métodos Estatísticos

Específicas

- Álgebra Linear Computacional
- Algoritmos em Grafos
- Aprendizado de Máquina
- Ciência de Redes
- Gerência de Dados em Larga Escala
- Mineração de Dados
- Mineração de Textos
- Otimização por Metaheurísticas
- Visualização de Dados
- Fundamentos de Sistemas Multimídia
- Aplicações de Robótica

Mercado de trabalho no Brasil e no mundo

- Nos EUA e em outros países as oportunidades são boas
- No Brasil existe uma defasagem
 - Em 2014 havia poucas ofertas
 - Hoje o número é razoável
- Oferta aparece com várias nomenclaturas diferentes
 - Machine Learning
 - Big Data
 - Business Intelligence, Data Analytics, Data Engineering



Não é uma panaceia

Estabilidade versus volatilidade

Expectativa versus realidade

É um nicho, portanto, deve ser encarado como um diferencial



CIÊNCIA DE DADOS: VISÃO GERAL E FORMAÇÃO



Eduardo Ogasawara
eogasawara@ieee.org
<https://eic.cefet-rj.br/~eogasawara>