



UFSC-CTC-INE



Programa de Pós-Graduação em  
Ciência da Computação - PPGCC

INE 410131

Gerência de Dados para Big Data



Ronaldo S. Mello  
2024/1



# Plano de Ensino da Disciplina

- Objetivo
- Requisitos
- Conteúdo
- Metodologia
- Avaliação
- Cronograma (Previsto) de Aulas
- Bibliografia

# Objetivo

Estudo e discussão de tecnologias para a gerência de dados para *Big Data*. A motivação para a pesquisa nesta área é o desafio relacionado à gestão eficiente de grandes volumes de dados heterogêneos produzidos por pessoas, dispositivos e sistemas em diversos domínios de aplicação.

# Requisitos

Recomenda-se que o aluno tenha conhecimentos sobre fundamentos de Banco de Dados (BD):

- bancos de dados relacionais
- SQL
- modelagem de dados

# Conteúdo

## 1. Introdução à Big Data

- 1.1 Evolução dos requisitos de gerência de dados e dos modelos de BD
- 1.2 Definição de Big Data
- 1.3 Desafios na gestão de Big Data: Os “Vs”
- 1.4 Exemplos de domínios de aplicação

## 2. BDs nas Nuvens

- 2.1 O paradigma de Computação nas Nuvens
- 2.2 Requisitos de gerência de dados nas nuvens
- 2.3 Teorema CAP, teorema PASELC e propriedades BASE
- 2.4 Categorias de BDs nas nuvens

## 3. BDs NoSQL

- 3.1 Definição e principais características
- 3.2 Modelos de dados
- 3.3 Projeto lógico
- 3.4 Principais SGBDs e seus recursos de gerenciamento de dados
- 3.5 Persistência poliglota de dados

## 4. BDs NewSQL

- 4.1 Definição e principais características
- 4.2 Modelos de dados
- 4.3 Principais SGBDs e seus recursos de gerenciamento de dados

## 5. BDs em Memória

- 5.1 Definição e principais características
- 5.2 Tecnologias envolvidas no seu desenvolvimento
- 5.3 Exemplos de SGBDs e seus recursos de gerenciamento de dados

## 6. Infraestruturas para Processamento de Big Data

- 6.1 Hadoop
- 6.2 Spark
- 6.3 Kafka

## 7. Tópicos em Gerência de Big Data

- 7.1 Integração de dados
- 7.2 Análise de dados (*Big Data Analytics*)
- 7.3 Tendências de pesquisa

# Metodologia

- Aulas e atividades de avaliação na modalidade não-presencial síncrona
- Leituras complementares e atividades de avaliação na modalidade não-presencial síncrona ou assíncrona
- Atividades de avaliação devem ser postadas no ambiente Moodle

# Avaliação

- Prova (peso: 3 – individual)
- Seminário (peso: 3 – grupo até 3 pessoas)
- Artigo (documento resultante de pesquisa técnico-científica – grupo até 3 pessoas) (peso: 3)
- Atividades práticas (peso: 1 – grupo até 3 pessoas)
- Envios fora do prazo para o meu e-mail não serão aceitos!



# Cronograma (Previsto) de Aulas

Data	Conteúdo
<b>12/3</b>	<b>Plano de ensino da disciplina</b> ; modelagem de dados
<b>19/3</b>	Introdução à Big Data
<b>26/3</b>	BDs nas nuvens; BDs NoSQL
<b>02/4</b>	BDs NoSQL
<b>09/4</b>	BDs NoSQL
<b>16/4</b>	BDs NoSQL
<b>23/4</b>	BDs NewSQL e BDs em memória
<b>30/4</b>	BDs NewSQL e BDs em memória
<b>07/5</b>	Persistência poliglota de dados
<b>14/5</b>	Infraestruturas para processamento de Big Data
<b>21/5</b>	Infraestruturas para processamento de Big Data
<b>28/5</b>	Desafios na gerência de Big Data
<b>04/6</b>	Desafios na gerência de Big Data
<b>11/6</b>	<b>PROVA</b>

Data	Conteúdo
<b>18/6</b>	<b>Seminários</b>
<b>25/6</b>	<b>Seminários</b>
<b>02/7</b>	<b>Seminários</b>
<b>09/7</b>	<b>Entrega do artigo</b>

# ERBD 2024



[Home](#) [Sobre a SBC](#) [Chamada de Trabalhos](#) [Inscrições](#) [Programação](#) [Equipe](#) [Local](#) \*

10 a 12 de Abril de 2024  
Farroupilha - RS

Realização



Execução



Parceria



## Tema

Inteligência Artificial Centrada em Dados

## Sobre o ERBD 2024

A Escola Regional de Banco de Dados (ERBD) é um evento anual, sem fins lucrativos, realizado pela Sociedade Brasileira de Computação (SBC). É um evento anual itinerante, que reúne especialistas da academia e da indústria da área de banco de dados para discussão e apresentação de trabalhos científicos, realização de palestras, minicursos, oficinas, painéis com debatedores e relatos de experiências e aplicações. Nessa edição o evento é executado pelo Instituto Federal do Rio Grande do Sul (IFRS) Campus Farroupilha.

<https://web.farroupilha.ifrs.edu.br/erbd24/>

# Bibliografia

1. Maheshwari, A. Big Data Essentials. Kindle Edition, 2016.
2. Bahga, A. and Madiseti, V. Big Data Science & Analytics. VPT Publisher, 1st Ed., 2016.
3. Sousa, F. R. C. et al. Gerenciamento de Dados em Nuvem: Conceitos, Sistemas e Desafios. Em: Tópicos em Sistemas Colaborativos, Interativos, Multimídia, Web e Bancos de Dados. Minicursos do XXV Simpósio Brasileiro de Banco de Dados (SBBD), 2010.
4. Monteiro, J. M.; Brayner, A. R. A.; Tavares, J. A. What Comes After NoSQL? NewSQL: A New Era of Challenges in DBMS Scalable Data Processing. Em: Tópicos em Gerenciamento de Dados e Informações. Minicursos do XXXI Simpósio Brasileiro de Banco de Dados (SBBD), 2016. p. 27-56.
5. Sadalage, P. J. and Fowler, M. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. 1<sup>a</sup> ed. Addison-Wesley, 2012.
6. Moniruzzaman, A. B. M. NewSQL: Towards next generation scalable RDBMS for OLTP for big data management. CoRR, abs/1411.7343.

# Bibliografia

7. Grolinger, K., Higashino, W. A., Tiwari, A., and Capretz, M. A. Data management in cloud environments: NoSQL and newSQL data stores. J. Cloud Computing, v.2, n.1, 2013.
8. Cattell, R. Scalable SQL and NoSQL data stores. SIGMOD Record, v.39, 2011. p. 12–27.
9. Dean, J. and Ghemawat, S. Mapreduce: simplified data processing on large clusters. Communications of ACM, v.51, n.1, 2008. p. 107–113.
10. Doshi, K. A., Zhong, T., Lu, Z., Tang, X., Lou, T., and Deng, G. (2013). Blending SQL and newSQL approaches: reference architectures for enterprise Big Data challenges. In: CYBERC, 2013. p. 163– 170.
11. Gilbert, S. and Lynch, N. A. Perspectives on the CAP theorem. IEEE Computer, v.45, n.2, 2012. p. 30–36.
12. Abadi, D. Consistency Tradeoffs in Modern Distributed Database System Design: CAP is Only Part of the History. IEEE Computer, v.45, n.2, 2012. p. 37-42.

# Bibliografia

13. Lima, C. Projeto Lógico de Bancos de Dados NoSQL Documento a Partir de Esquemas Conceituais Entidade-Relacionamento Estendido (EER). Dissertação de Mestrado: Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Santa Catarina (PPGCC-UFSC), 2016.
14. Marz, N. and Warren, J. Big Data: Principles and Best Practices of Scalable Real-Time Data Systems. Manning Publishing Company, 2015.
15. Dong, X. L. and Srivastava, D. Big Data Integration. Morgan & Claypool Publishers, 2015.
16. Bordin, M. V. et al. Trabalhando com Big Data em Tempo Real. Minicursos do XVI Escola Regional de Alto Desempenho do Rio Grande do Sul (ERAD/RS), 2016.



# Web Sites Interessantes

- [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data)
- <http://nosql-database.org/>
- <http://hadoop.apache.org/>
- [https://en.wikipedia.org/wiki/Data\\_analysis](https://en.wikipedia.org/wiki/Data_analysis)
- <https://en.wikipedia.org/wiki/NewSQL>
- <https://db-engines.com/>

## *Busca por artigos acadêmicos:*

- <http://scholar.google.com.br/> (Google acadêmico)
- <http://www.informatik.uni-trier.de/~ley/db/> (DBLP)
- <http://citeseer.ist.psu.edu/> (Citeseer)
- <http://dl.acm.org/> (ACM digital library)
- <http://ieeexplore.ieee.org/> (IEEE digital library)
- <https://www.scopus.com> (Elsevier Scopus)