



Universidade Federal de Santa Catarina  
Centro Tecnológico  
Departamento de Informática e Estatística  
Programa de Pós-Graduação em Ciência da Computação

## Plano de Ensino

### 1) Identificação:

**Disciplina:** INE410131– Gerência de Dados para *Big Data*

**Carga horária:** 45 horas/aula

**Professor:** Ronaldo dos Santos Mello

### 2) Requisitos:

Esta disciplina requer que o aluno tenha conhecimentos sobre fundamentos de Banco de Dados, principalmente o modelo relacional, SQL, modelagem de dados e transações.

### 3) Ementa:

Evolução dos requisitos de gerenciamento de dados e dos modelos de bancos de dados. Big Data: definição, características e exemplos de domínios de aplicação; Bancos de dados nas nuvens e suas categorias; Bancos de dados NoSQL; Bancos de dados NewSQL; Bancos de dados em memória; Infraestruturas para processamento de Big Data; Desafios na gerência de Big Data: integração, armazenamento, análise de dados e suporte a tempo real.

### 4) Objetivos:

**Geral:** Estudo e discussão de soluções para a gerência de dados para Big Data. A motivação para a pesquisa nesta área é o desafio para lidar eficientemente com volumes massivos de dados heterogêneos produzidos por pessoas, dispositivos e sistemas em diversos domínios de aplicação.

#### **Específicos:**

- Caracterizar Big Data;
- Estudar os diferentes tipos de bancos de dados voltados ao gerenciamento de Big Data;
- Estudar as principais tecnologias disponíveis para o processamento de Big Data;
- Entender as principais problemáticas associadas à gerência de Big Data e soluções existentes;
- Pesquisar, sumarizar e apresentar alguma temática relevante de pesquisa na área.

### 5) Conteúdo Programático

#### 1. Introdução à Big Data

1.1 Evolução dos requisitos de gerenciamento de dados e dos modelos de bancos de dados

1.2 Definição de Big Data

1.3 Características: os 5 “Vs”

1.4 Principais domínios de aplicação

#### 2. Bancos de Dados 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

2.3 Categorias de bancos de dados nas nuvens

#### 3. Bancos de Dados 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
- 4. Bancos de Dados NewSQL
  - 4.1 Definição e principais características
  - 4.2 Categorias
  - 4.3 Principais SGBDs e seus recursos de gerenciamento de dados
- 5. Bancos de Dados em Memória
  - 5.1 Definição e principais características
  - 5.2 Tecnologias envolvidas no seu desenvolvimento
  - 5.3 Principais SGBDs e seus recursos de gerenciamento de dados
- 6. Infraestruturas para Processamento Distribuído de Big Data
  - 6.1 Hadoop
  - 6.2 Spark
  - 6.3 Kafka
- 7. Desafios na Gerência de Big Data
  - 7.1 Armazenamento
  - 7.2 Integração
  - 7.3 Análise (*Big Data Analytics*)
  - 7.4 Suporte a tempo real (*Fast Data*)

## 6) Metodologia

1. Aulas expositivas
2. Leituras complementares em aula ou extraclasse
3. Participação em aula
  - 3.1 Exercícios propostos (baseado no conteúdo apresentado em aula ou em artigos acadêmicos)
  - 3.2 Seminários

## 7) Avaliação

A avaliação da disciplina compreende as seguintes atividades:

1. Prova (peso: 3);
2. Seminário (peso: 3);
3. Trabalho escrito (artigo ou outro tipo de documento resultante de pesquisa técnico-científica) (peso: 3);
4. Exercícios propostos (peso: 1).

## 8) Cronograma

- 1ª semana: apresentação da disciplina e introdução a Big Data;
- 2ª semana: bancos de dados nas nuvens e bancos de dados NoSQL;
- 3ª semana: bancos de dados NoSQL;
- 4ª semana: bancos de dados NoSQL;
- 5ª semana: bancos de dados NewSQL e bancos de dados em memória;
- 6ª semana: bancos de dados NewSQL e bancos de dados em memória;
- 7ª semana: infraestruturas para processamento de Big Data;
- 8ª semana: infraestruturas para processamento de Big Data e definição de temas para seminários e trabalho escrito;
- 9ª semana: desafios na gerência de Big Data;
- 10ª semana: desafios na gerência de Big Data e PROVA;
- 11ª a 14ª semanas: apresentação de seminários;
- 14ª semana: entrega do trabalho escrito.

## 9) 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 (SBBDD), 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 (SBBDD), 2016. p. 27-56.
5. Sadalage, P. J. and Fowler, M. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. 1ª ed. Addison-Wesley, 2012.
6. Moniruzzaman, A. B. M. NewSQL: Towards next generation scalable RDBMS for online transaction processing (OLTP) for big data management. CoRR, abs/1411.7343.
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: International Conference on CyberEnabled Distributed Computing and Knowledge Discovery (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.
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.