

---

**Título:** BIG-065 - BIG DATA - Processamento e Análise de Dados em Larga Escala

**Área de Concentração:** Engenharia de Software

**Obrigatória:** Não

**Criação:** 20/10/2016

**Alteração:**

**Carga Horária:** 52

**Crédito:** 5

**Responsável:** Fabio Silva Lopes

**Observações:** Durante o curso, será desenvolvido de um projeto de Big Data e será elaborada uma resenha de 10 páginas aprofundando um dos temas discutidos em aula. As aulas expositivas e a discussão de artigos selecionados subsidiarão a elaboração destes trabalhos. A resenha deve apresentar uma proposta de evolução para o limiar dos temas de pesquisa apresentados. Os projetos proverão a abordagem prática complementar ao conteúdo programático do curso utilizando ferramentas disponíveis no mercado.

**Objetivo:** Apresentar conceitos, mecanismos especializados e aplicações, com base em Plataforma Big Data, para processamento e análise de dados em larga escala. O aluno terá a oportunidade de trabalhar na investigação dos limites desta área de pesquisa, por meio da discussão de artigos científicos e práticas de laboratório.

**Justificativa:** *Implicações no âmbito da pesquisa:*

A pesquisa nesta área é relevante e presente, com foco em demandas de larga escala, a considerar que o volume de dados passíveis de coleta, processamento e análise, vêm aumentando sobremaneira e exigindo novos patamares de processamento para atender aplicações que requerem desempenho e disponibilidade. De modo complementar, técnicas estatísticas e de mineração de dados, se aliam na descoberta de padrões que podem indicar tendências, estabelecer associações, ou classificar conjuntos de modo a permitir um melhor entendimento sobre objetos de estudo.

*Aplicações práticas:*

Tecnologias associadas a Big Data e análise de dados são essenciais para subsidiar novos insumos com potencial de alavancar negócios. Dados são considerados “assets” ou bens de uma organização. Sua aquisição se dá por meio da coleta em fluxo, intra e/ou extra organização, a partir de várias fontes. São tratados e armazenados em plataformas apropriadas para o processamento, análise e disponibilização em larga escala. O domínio e a gestão destes recursos são essenciais para agregar valor em diversas modalidades de negócios ou científicas, que buscam a identificação de novos padrões ainda não observados nestas coleções de dados.

**Ementa:** Conceito de organizações e sistemas orientadas a dados; Arquiteturas e plataformas orientadas a dados; Requisitos de disponibilidade, desempenho e escala; Coleta em fluxo; Persistência NoSQL; Processamento em cluster; Análise e visualização de dados, Qualidade e Governança de dados. Desenvolvimento de uma aplicação prática com aplicação dos conteúdos abordados a um contexto proposto.

---

Forma de Avaliação: AVALIAÇÃO 1: PROJETO EM GRUPO  
AVALIAÇÃO 2: MONOGRAFIA EM GRUPO  
AVALIAÇÃO 3: AVALIAÇÃO INDIVIDUAL FINAL  
MÉDIA DE (AVALIAÇÃO 1 + AVALIAÇÃO 2 + AVALIAÇÃO 3)  $\geq 6.0$

Material Utilizado: Recursos de Projeção, Portal de periódicos e Moodle

Metodologia: Recursos de Projeção, Portal de periódicos e Moodle

Conhecimentos Prévio: Conhecimentos de bancos de dados relacionais e estatística descritiva.

Bibliografia Básica: Anderson, C. **Creating a Data-Driven Organization: Practical Advice from the Trenches**. O'Reilly, 2015.  
DAMA International. **The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK)**. Technics Publications, 2010.  
Larose, D.T.; Larose, C. **Data Mining and Predictive Analytics**. 2nd ed. Wiley, 2015.  
Mohanty, H.; Bhuyan, P.; Chenthati, D. (ed). **Big Data A Primer**. Springer, 2015.  
Sadalage, P.; Fowler, M. **NoSQL Distilled: A brief Guide to the Emerging World of Polyglot Persistence**. Pearson, 2013.  
Silva, L.A; Peres, S.M; Boscarioli, C. **Introdução a Mineração de dados: Com aplicações em R**. Elsevier, 2016.  
Somasundaram, G.; Shrivastava, Alok; EMC. **Armazenamento e Gerenciamento de Informações: Como armazenar, gerenciar e proteger informações digitais**. Bookman, 2011.  
Schutt, R.; O'Neil, C. **Doing Data Science: Straight Talk from the Frontline**. O'Reilly, 2013.  
Tan, P.N.; Steinbach, M.; Kumar, V. **Introduction to Datamining**. Addison Wesley, 2006.  
Zicopoulos, P., et al. **Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data**. McGraw-Hill, 2012.

Bibliografia Complementar: Cattell, R. Scalable SQL and NoSQL data stores. **ACM SIGMOD Record**, 39(4), 12-27, 2011.  
Chen, H., Chiang, R. H., & Storey, V. C. Business Intelligence and Analytics: From Big Data to Big Impact. **MIS quarterly**, 36(4), 1165-1188, 2012.  
Fan, W., & Bifet, A. Mining big data: current status, and forecast to the future. **ACM SIGKDD Explorations Newsletter**, 14(2), 1-5, 2013.  
Han, J., Haihong, E., Le, G., & Du, J. Survey on NoSQL database. In Pervasive computing and applications (ICPCA), 2011 **6th IEEE international conference on** (pp. 363-366), 2011.  
Normas da Família ISO/IEC 27000. Sistema de Gestão de Segurança da Informação.  
LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. Big data, analytics and the path from insights to value. **MIT Sloan Management Review**, 21, 2013.  
Madden, S. From databases to big data. **IEEE Internet Computing**, 16(3), 0004-6, 2012.

---

Programa da Disciplina: **CRONOGRAMA DE AULAS**

**AULA CONTEÚDO**

**AULA 1**

CONCEITOS DE ORGANIZAÇÕES E SISTEMAS DATA-DRIVEN (ORGANIZAÇÕES E SISTEMAS ORIENTADOS A DADOS, DIFERENTES TIPOS, FONTES DE DADOS E DEMANDAS, PESQUISAS E DESAFIOS DA ÁREA DE PROCESSAMENTO E ANÁLISE DE DADOS EM LARGA ESCALA) (DEFINIÇÃO DE GRUPOS DE TRABALHO E TEMAS PARA ARTIGOS).

**AULA 2**

ARQUITETURAS E PLATAFORMAS ORIENTADAS A DADOS (DISTRIBUTED FILE SYSTEM E OPÇÕES HÍBRIDAS, FERRAMENTAS DE APOIO).

**AULA 3**

DISPONIBILIDADE, DESEMPENHO E ESCALA, PROCESSAMENTO PARALELO, PERSISTÊNCIA NoSQL (ATIV. PRÁTICA, INSTALAÇÃO E MANIPULAÇÃO DE ARQUIVOS).

**AULA 4**

COLETA DE DADOS (SENSORIAMENTO, CONECTORES, FERRAMENTAS E CONECTORES PARA COLETA EM FLUXO).

**AULA 5**

ANÁLISE DE DADOS (ESTATÍSTICA DESCRITIVA E MODELOS DE REGRESSÃO, FERRAMENTAS PARA ANÁLISE – LABORATÓRIO)

**AULA 6**

ANÁLISE DE DADOS (MINERAÇÃO DE DADOS - CLASSIFICAÇÃO, FERRAMENTAS PARA ANÁLISE – LABORATÓRIO)

**AULA 7**

ANÁLISE DE DADOS (MINERAÇÃO DE DADOS - ASSOCIAÇÃO, FERRAMENTAS PARA ANÁLISE – LABORATÓRIO)

**AULA 8**

ANÁLISE DE DADOS (MINERAÇÃO DE DADOS - CLUSTER, FERRAMENTAS PARA ANÁLISE – LABORATÓRIO)

**AULA 9**

VISUALIZAÇÃO DE DADOS (TABELAS, GRÁFICOS, MAPAS, NUVEM, OUTRAS, FERRAMENTAS PARA VISUALIZAÇÃO – LABORATÓRIO)

**AULA 10**

QUALIDADE E GOVERNANÇA DE DADOS (MASTER DATA MANAGEMENT e METADADOS, OS DADOS EM FRAMEWORKS DE MATURIDADE, VISÃO DA INFORMAÇÃO EM RM-ODP, DM-BOOK)

**AULA 11**

APRESENTAÇÃO DOS ARTIGOS

**AULA 12**

APRESENTAÇÃO DOS PROJETOS

**AULA 13**

AVALIAÇÃO