



MINISTÉRIO DA EDUCAÇÃO  
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA  
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA  
PRÓ-REITORIA DE ENSINO

PLANO DE DISCIPLINA			
IDENTIFICAÇÃO			
CAMPUS: JOÃO PESSOA			
CURSO: BACHARELADO EM ENGENHARIA DE SOFTWARE			
DISCIPLINA: <b>Engenharia de Dados</b>			
CÓDIGO DA DISCIPLINA: ES65			
PRÉ-REQUISITO: Banco de Dados II, Probabilidade e Estatística			
UNIDADE CURRICULAR: Obrigatória [x] Optativa [ ] Eletiva [ ]			SEMESTRE:
CARGA HORÁRIA			
TEÓRICA: 37 h	PRÁTICA: 30 h	EaD <sup>1</sup> : Não	EXTENSÃO: -----
CARGA HORÁRIA SEMANAL: 4 horas-aula			
CARGA HORÁRIA TOTAL: 67 h			
DOCENTE RESPONSÁVEL: Diego Ernesto Rosa Pessoa			

EMENTA
--------

**Fontes e tipos de dados. Pipelines de dados para desenvolvimento de aplicações baseadas em dados. Modelos de bancos de dados não relacionais (NoSQL): características, linguagem de definição, linguagem de manipulação e aplicabilidade. Persistência poliglota. Coleta de dados em diferentes fontes. Preparação e pré-processamento de dados. Carga de dados em sistema de banco de dados não relacional. Categorias de Análises de Dados. Aplicação de análise de dados.**

OBJETIVOS
-----------

**Geral:**

- Compreender o processo de Engenharia de Dados por meio da implementação de pipelines de dados, visando realizar a coleta, preparação, pré-processamento e persistência de dados para aplicações dirigidas a dados.

**Específicos:**

- Conhecer os conceitos e práticas da Engenharia de Dados, compreendendo seu papel fundamental na análise e utilização de dados em diferentes contextos.
- Implementar pipelines de dados, incluindo a definição de etapas de coleta, transformação e carga de dados, a fim de garantir a integridade e a qualidade dos dados para uso posterior.
- Aprender técnicas e ferramentas de coleta de dados, explorando diferentes fontes e tipos de dados relevantes para aplicações orientadas a dados.
- Adquirir habilidades de preparação e pré-processamento de dados, compreendendo a importância da limpeza, transformação e agregação de dados para garantir sua qualidade e consistência.
- Utilizar modelos de bancos de dados não relacionais como ferramenta para persistência de dados coletados e processados no pipeline de dados.

<sup>1</sup> Para a oferta de disciplinas na modalidade à distância, integral ou parcial, desde que não ultrapasse 20% (vinte por cento) da carga horária total do curso, observar o cumprimento da Portaria MEC no 1.134, de 10 de outubro de 2016.

- Integrar conhecimentos teóricos e práticos, utilizando conjuntos de dados reais.

## BIBLIOGRAFIA

### **Bibliografia Básica:**

1. AMARAL, F. Introdução à Ciência de Dados, Mineração de Dados e Big Data. Ed. Alta Books, 2016.
2. BOAGLIO, F. MongoDB - Construa novas aplicações com novas tecnologias. São Paulo: Casa do Código, 2015.
3. DENSMORE, J. Data Pipelines Pocket Reference: Moving and Processing Data for Analytics. 1 ed. O'Reilly Media, Inc. 2021.
4. HOUSLEY, M., REIS, J. Fundamentals of Data Engineering. O'Reilly Media, Inc. 2022.
5. SADALAGE, P. J. NoSQL: um guia conciso para o mundo emergente da persistência poliglota. São Paulo: Novatec, 2013. 220p. ISBN: 9788575223383

### **Bibliografia Complementar:**

1. BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca. Editora Bookman; 2ª edição, 2013. ISBN : 978-8582600481
2. DAVID, H. Introdução ao MongoDB. São Paulo: Novatec, 2015. 167 p. ISBN: 9788575224229
3. Documentação do MongoDB. Disponível em <https://docs.mongodb.com/>. Último acesso em 11/03/2021.
4. DOAN, A.; HALEVY, A. e IVES, Z. Principles of Data Integration.1. ed. [S.l.]: Morgan Kaufmann, 2012.
5. DONG, X. L.; SRIVASTAVA, D. Big Data Integration. [S.l.]: Morgan & Claypool. 2015.
6. ELMASRI, Ramez; Navathe, Shamkant B. Sistemas de Banco de Dados. Editora Pearson/Addison Wesley Pub Co Inc. 7a Edição, 2019. ISBN: 978-8543025001.
7. KELLEHER, J.; TIERNEY, B. Data Science. The MIT Press. 2018. ISBN: 9780262347037.
8. KIMBAL, Ralph; MERZ, Richard. Data Webhouse: construindo o data warehouse para a web. Rio de Janeiro: Campus, 2000. 365 p
9. Padrões e recomendações do W3C. Disponível em <<https://www.w3.org/TR/dwbp/>>. Último acesso em: 18 fevereiro de 2021.
10. MACHADO, F. N. R, Big data : o futuro dos dados e aplicações. São Paulo : Érica, 2018. - 224 p. ISBN: 9788536527000
11. MCKINNEY, W. Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython. Editora : Novatec Editora; 1ª edição (10 maio 2019).