



## Plano de Disciplina

Identificação				
CURSO Mestrado Profissional em Tecnologia da Informação			CAMPUS João Pessoa	
DISCIPLINA Integração de Dados			CÓDIGO DA DISCIPLINA 87407	
PRÉ-REQUISITO Não há pré-requisito				
UNIDADE CURRICULAR (OBRIGATÓRIA, OPTATIVA, ELETIVA) Eletiva			SEMESTRE 2024.2	
DOCENTE RESPONSÁVEL Diego Ernesto Rosa Pessoa				
Carga Horária				
TEÓRICA 24	PRÁTICA 24	EaD <sup>1</sup> 0	CARGA HORÁRIA SEMANAL: 4	CARGA HORÁRIA TOTAL 60

## Ementa

Alinhamento de Esquemas, resolução de Entidades e Fusão de Dados. Fontes de dados: tipos, descrição, heterogeneidade, autonomia. Big Data: definição, características e exemplos de domínios de aplicação; Big Data como Ambiente de integração de dados (infraestrutura e tecnologia para processamento de Big Data); Visualização de Dados; Modelos de dados para Integração; Desafios em Big Data.

## Objetivos

Aprender sobre o problema da integração de dados. Definir esquema de mapeamento e visões integradas. Descrever, conceituar e caracterizar Big Data. Identificar e caracterizar aplicações no cenário de Big Data. Explicar os requisitos para modelos de programação para Big Data. Apresentar o ecossistema Spark. Conhecer a importância da modelagem e gerenciamento de Big Data. Descrever Ciência de Dados: conceitos, processo e uso de ferramentas para análise de dados.

## Conteúdo Programático

Tópico	Carga-horária
Apresentação da Disciplina – visão geral em Integração de Dados	4
Introdução a Big Data (definição, características, domínios de aplicação). Modelos de Programação para Big Data.	8
Fontes de dados, heterogeneidade, sistemas de arquivos distribuídos	8
Sistemas de Gerenciamento de fluxos de dados. Ingestão, Armazenamento, Qualidade de Dados. Modelagem de Big Data.	12
Integração de dados em Big Data. Processamento de Integração de big data. Recuperando Big Data. Integrando Big Data. Análise de Big Data.	12
Desafios de Pesquisa em Integração de Dados.	12
Apresentação de Projetos e Relatório Técnico	4
<b>Total</b>	<b>60</b>



## **Metodologia de Ensino**

Aulas teóricas e práticas.

Análises e discussão sobre cases, pesquisas e soluções no escopo da disciplina.

Estudos complementares com artigos e ferramentas.

Acompanhamento de atividades discentes através de plataforma de apoio a aprendizagem (Google Classroom)

Planejamento e execução de projeto de fluxo de execução (pipeline) para integração de dados

Reuniões de acompanhamento dos projetos para cada time de desenvolvimento e entrega de milestones em datas pré-estabelecidas.

## **Recursos Didáticos**

Classroom, ferramentas de software open source, periódicos, artigos, acesso à Internet.

## **Critérios de Avaliação**

A avaliação será realizada com base em duas etapas: (i) Projeto de aplicação com demonstração de aplicação utilizando pipeline para integração de dados; (ii) Relatório técnico com definições e resultados associados ao projeto da aplicação, com apresentação oral e escrita. Todas as atividades terão etapas e deadlines definidos a serem cumpridos.

## **Bibliografia**

1. BORGMAN, C. Big data, Little Data, No Data. MIT Press, 2015.
2. KRISHNAN, Krish. Building Big Data Applications. 1st. ed., Elsevier 2019. 2. KNAFLIC, Cole Nussbaumer. Storytelling with Data: Let's Practice!. 1st ed. Wiley 2019.
3. KLEPPMANN, Martin. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. 1st ed., O'Reilly Media 2017.
4. MARZ, Nathan; WARREN, James. Big Data: Principles and Best Practices Scalable Realtime Data Systems. 1st ed., Manning Publications 2015.
5. DAMJI, Jules S.; WENIG, Brooke; DAS, Tathagata; LEE, Denny. Learning Spark: Lightning-Fast Big Data Analysis. 2nd ed., O'Reilly Media 2020.
6. MACHADO, Nery Felipe Rodrigues. Big Data: O Futuro dos Dados e Aplicações. 1a ed. Érica 2018.
7. DONG, X. L.; SRIVASTAVA, D. Big Data Integration. [S.I.]: Morgan & Claypool. 2015.
8. Densmore, James; James Densmore - Data Pipelines Pocket Reference, Moving and Processing Data for Analytics-O'Reilly Media, Inc, USA, 2021
9. Bas Harenslak and Julian de Ruiter. Data Pipelines with Apache Airflow. Manning Publications 2021
10. Mamta Mittal, Valentina E. Balas, Lalit Mohan Goyal, Raghvendra Kumar. Big Data Processing Using Spark in Cloud., Springer 2019.
11. Jonathan Rioux. Data Analysis with Python and PySpark. Manning Publications 2021



## **Observações**

(Nenhuma)