

CURSO: Ciência de Dados e Inteligência Artificial – 1º semestre de 2023

DISCIPLINA: **Computação Escalável**

PROFESSOR(ES): Thiago Pinheiro de Araújo

CARGA HORÁRIA: 60h

PRÉ-REQUISITO: Banco de Dados; Técnicas e Algoritmos em Ciência de Dados

PLANO DE ENSINO

1. Ementa

Programação Concorrente e Paralela; Sistemas Operacionais: processos, escalonadores e memória; Processos versus Threads; Memória Compartilhada; Mecanismos de Sincronização; Problemas Clássicos de Concorrência; Paralelismo e Concorrência em Python; Mecanismos de Comunicação através de Mensagens; Map Reduce, Hadoop, Spark; Introdução à Computação em Nuvem;

2. Objetivos da disciplina

Esta disciplina tem como objetivo geral introduzir conceitos da área de Sistemas Distribuídos, especificamente, teorias, técnicas, e ferramentas para construir aplicações de processamento de dados utilizando soluções de paralelismo e distribuição.

3. Procedimentos de ensino (metodologia)

Aulas teóricas ministradas através de slides sobre os temas; e aulas práticas ministradas através de aplicações, estudos de caso e exercícios baseados em problemas reais. As aulas serão complementadas com projetos e leitura dirigida.

4. Conteúdo programático detalhado

Datas	Tópico	Atividades
14/02	Conceitos gerais, motivação e aplicações	Apresentação da disciplina
16/02	Concorrência e Paralelismo	Apresentação da teoria
29/02	Sistemas operacionais: processos, escalonadores e memória	Apresentação da teoria
02/03	Processos versus Threads	Apresentação da teoria
07/03	Processos versus Threads	Prática através de exercícios
09/03	Trabalho 1	Apresentação e discussão do enunciado
14/03	Acesso concorrente à dados, seções críticas e mecanismos de controle	Apresentação da teoria
16/03	Acesso concorrente à dados, seções críticas e mecanismos de controle	Apresentação da teoria
21/03	Acesso concorrente à dados, seções críticas	Prática através de

	e mecanismos de controle	exercícios
23/03	Acesso concorrente à dados, seções críticas e mecanismos de controle	Prática através de exercícios
28/03	Problemas clássicos de concorrência	Apresentação da teoria
30/03	Problemas clássicos de concorrência	Prática através de exercícios
04/04	Paralelismo e concorrência em Python	Apresentação da teoria
11/04	Paralelismo e concorrência em Python	Apresentação da teoria
13/04	Paralelismo e concorrência em Python	Prática através de exercícios
25/04	Comunicação através de mensagens (Introdução)	Apresentação da teoria
27/04	Comunicação através de mensagens (MPI)	Apresentação da teoria
02/05	Comunicação através de mensagens (MPI)	Prática através de exercícios
04/05	Comunicação através de mensagens (RPC)	Apresentação da teoria
09/05	Comunicação através de mensagens (RPC)	Prática através de exercícios
11/05	Trabalho 2	Apresentação e discussão do enunciado
16/05	Comunicação através de mensagens (Publish-subscribe)	Apresentação da teoria
18/05	Comunicação através de mensagens (Publish-subscribe)	Prática através de exercícios
23/05	Map Reduce, Hadoop e Spark	Apresentação da teoria
25/05	Introdução à PySpark	Apresentação da teoria
30/05	Introdução à PySpark	Prática através de exercícios
01/06	Computação em nuvem	Apresentação da teoria
06/06	Computação em nuvem	Prática através de exercícios
13/06	Pipelines ETLs baseado em Spark	Apresentação da teoria
15/06	Trabalho 2	Apresentação dos grupos e discussão

5. Procedimentos de avaliação

Os alunos serão avaliados através de trabalhos em grupo.

6. Bibliografia Obrigatória

- Peter Pacheco. An Introduction to Parallel Programming. (Morgan Kaufmann) (ISBN 978-0123742605)
- Bill Chambers and Matei Zaharia. Spark: The Definitive Guide: Big Data Processing Made Simple. (O'Reilly Media) (ISBN 978-1491912218)
- Holden, Andy, and Patrick. Learning Spark: Lightning-Fast Big Data Analysis. O'Reilly. 2015.

7. Bibliografia Complementar

- Jimmy Lin and Chris Dyer. Data-Intensive Text Processing with MapReduce. (Morgan & Claypool) (ISBN 978-1608453429)
- Tom White. Hadoop: The Definitive Guide, Fourth Edition, by (O'Reilly) (ISBN 978-1-4919-0163-2)
- Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. 2010. A view of cloud computing. Commun. ACM 53, 4 (April 2010), 50-58. DOI: <https://doi.org/10.1145/1721654.1721672>
- Kai Hwang, Jack Dongarra & Geoffrey C. Fox. Distributed and Cloud Computing: Clusters, Grids, Clouds, and the Future Internet. <<https://www.elsevier.com/books/distributed-and-cloud-computing/hwang/978-0-12-385880-1>>
- Li, Kuan-Ching. Handbook of Research on Scalable Computing Technologies, Volumes 1-2. IGI Global

8. Minicurrículo do(s) Professor(s)

Possui graduação em Engenharia da Computação (2007) pela PUC-Rio, mestrado (2009) e doutorado (2014) em Ciência da Computação também pela PUC-Rio. Suas principais áreas de atuação são Engenharia de Software e Sistemas Distribuídos, com foco em: qualidade de software, sistemas auto-adaptáveis e sistemas orientados à recuperação. Thiago também é um empreendedor serial, desenvolvendo startups e projetos de inovação nos últimos 12+ anos.

9. Link para o Currículo Lattes

<http://lattes.cnpq.br/0851924048933884>