

CONSULTORIA ESPECIALIZADA: SISTEMAS DISTRIBUÍDOS

APRESENTAÇÃO DA UNIDADE CURRICULAR

PROF. CARLOS PAES.



PUC-SP

OBJETIVOS



- Proporcionar capacidades e habilidades para que o estudante:
 - Conheça e compreenda os conceitos básicos de sistemas computacionais distribuídos e paralelos
 - Conheça os principais paradigmas de comunicação em sistemas distribuídos e as principais questões relacionadas ao projeto de sistemas distribuídos
 - Aplicar esses conhecimentos em plataformas computacionais capazes de processarem grandes volumes de dados (big data) de forma distribuída e paralela
 - Analisar e avaliar programas desenvolvidos para um ambiente de execução constituído por diversas unidades de processamento operando em paralelo

CONTEÚDO PROGRAMÁTICO

- Conceitos básicos de sistemas distribuídos
- Conceitos básicos de programação paralela
- Aplicações distribuídas e paralelas
- Características básicas da Programação Distribuída
- Arquitetura de Sistemas Distribuídos
- Paradigmas de Comunicação Distribuída
- Classes de linguagens para a Programação Distribuída
- Plataforma de processamento de grande volumes de dados (big data) de forma distribuída e paralela: MapReduce. Spark. Resilient Distributed Datasets (RDD)

AVALIAÇÃO



- Atividades:
 - Experimentos no práticos
 - Exercícios em sala de aula
- Projeto Integrador
- Seminários Arquitetura de Sistemas Distribuídos e Tecnologias

PRÉ-REQUISITOS

- Conhecimento da linguagem de programação Python
- Algoritmos e estrutura de dados
- Conhecimento de organização de computadores (OC) e Arquitetura de Computadores (ARQ)
- Sistemas Operacionais e Programação Concorrente
- Redes de Computadores
- Banco de Dados (Estruturado e Não Estruturado)



DINÂMICA DAS AULAS

- Sala de aula invertida e Aprendizado baseado em projetos
- Exposições do professor no inicio das aulas
- Discussões e trabalho em grupo
- Experimentos práticos
- Exercícios individuais ou em grupo
- Projeto de implementação



LABORATÓRIO DE SD (PRÁTICA)



- Programação na linguagem Python
- Sistema Operacional Linux (Unix like)
- Bibliotecas Thread Python e CUDA
- Prática de programação paralela e distribuída
- Problemas de programação paralela e distribuída
- POCs de Tecnologias para desenvolvimento de arquiteturas distribuídas e plataforma de processamento de grande volumes de dados (big data) (Spark e Hadoop)

AMBIENTE VIRTUAL DE APRENDIZAGEM



- Equipe no Microsoft Teams
- Repositório de arquivos usado para disponibilizar material
- Github para publicação do projeto (comunidade de desenvolvedores de código aberto)
- Chat para nossa interação durante o semestre.
- Experimentos e projeto disponibilizados na guia Trabalhos
- Entregas realizadas pelo Teams ou Github → orientações fornecidas pelo professor
- Critérios de avaliação disponível para cada experimento e projeto

REGRAS DO JOGO



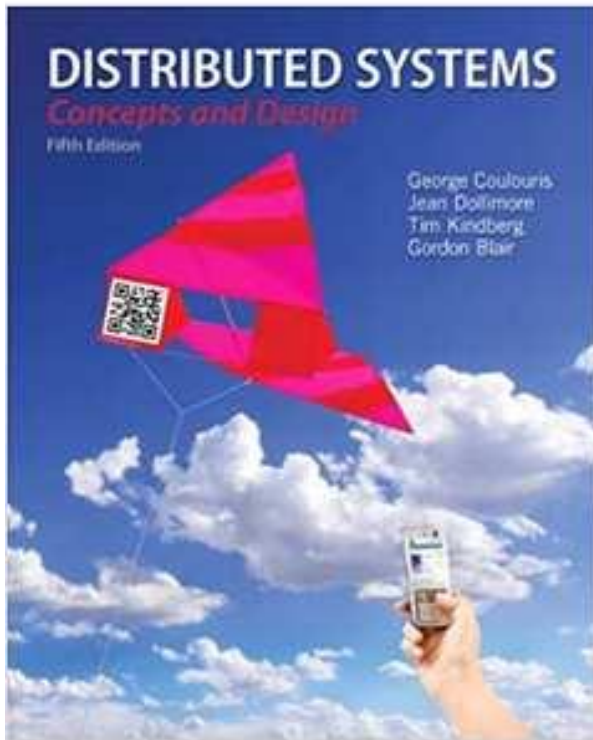
- Os experimentos e projetos deverão ser entregues por meio do Teams e dentro do prazo estabelecido pelo professor
- Não serão aceitos entregas por e-mail e chat do Teams.
- Não haverá prorrogação de prazo
- Não haverá, em hipótese alguma, substituição ou reposição para atividade perdida por não comparecimento à aula na qual foi realizada a atividade
- Trabalhos copiados ou com indícios de cópia serão zerados

COMPORTAMENTO DURANTE AS AULAS PRESENCIAIS

- Silêncio durante as aulas
- Não usar o celular
- Participe ativamente das aulas
- Chegar no horário
- Tire dúvidas durante as aulas

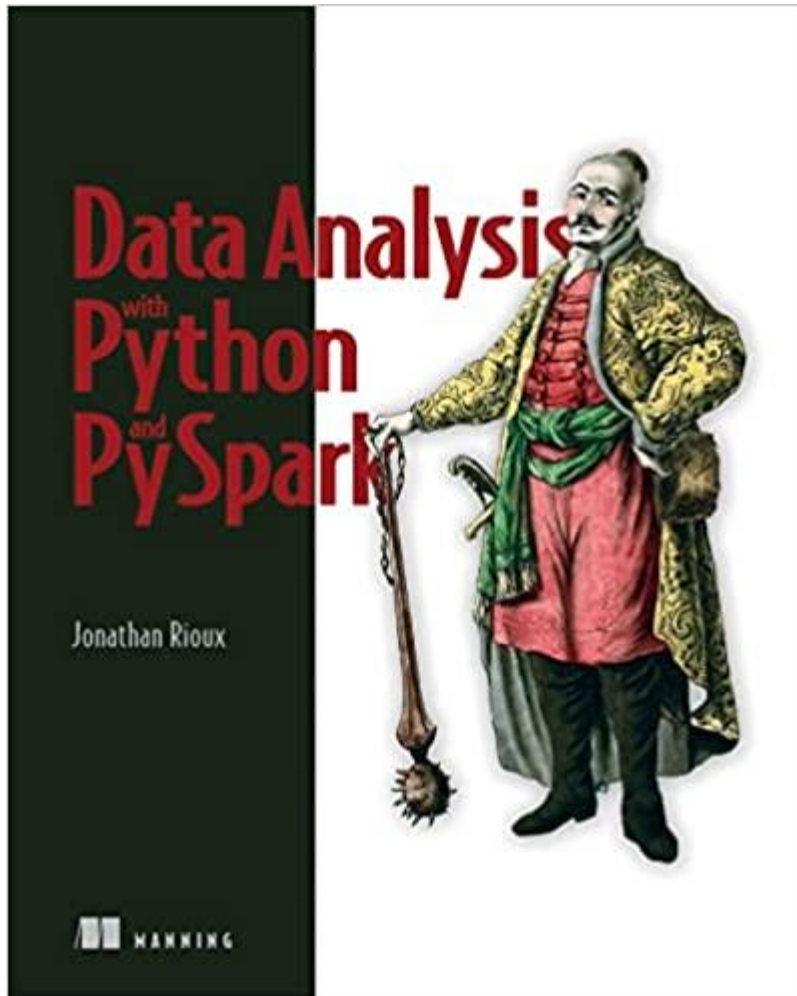


REFERENCIAS BIBLIOGRÁFICAS

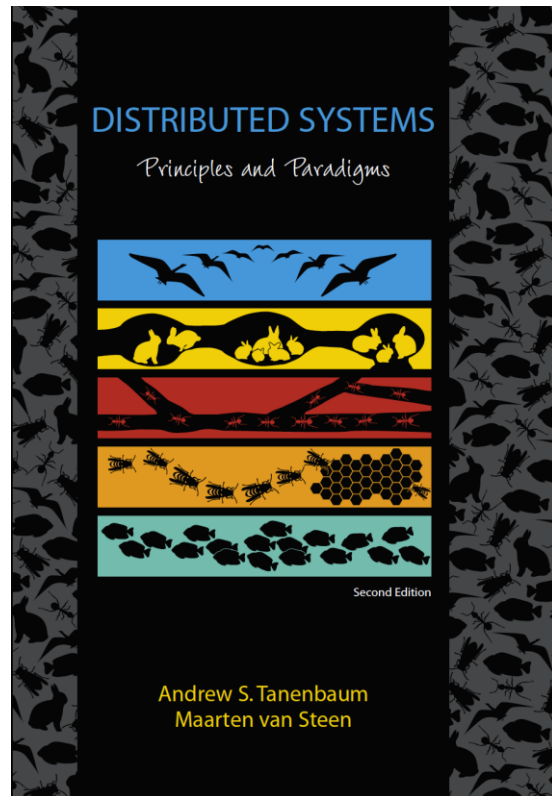


- Dollimore, J., T. Kindberg and G. Coulouris, Distributed Systems: Concepts and Design, 5.ed., Addison Wesley, 2011

REFERENCIAS BIBLIOGRÁFICAS



- RIOUX, J. Data Analysis with Python and Pyspark. [S.l.]: Manning Publications, 2022



- TANENBAUM, A. S.; VAN STEEN, M. Distributed systems: principles and paradigms. 2. ed

MOTIVAÇÕES



- Conceitos de Sistemas Distribuídos (SD) são relevantes em toda ciência da computação e ciência da dados/IA
- Um bom entendimento dos conceitos de SD fornece uma excelente base para a construção de sistemas para processamento de grande volumes de dados e de alta performance (HPC) → uso de computação distribuída e paralela
- Entender o SD é fundamental para um entendimento profundo de um sistema de computação orientado a dados e inteligentes (IA).



OBRIGADO

CARLOSP@PUCSP.BR