



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO

FACULDADE	Faculdade de Estudos Interdisciplinares
DEPARTAMENTO	
CURSO	Ciência de Dados e Inteligência Artificial
NOME DA DISCIPLINA	Consultoria Especializada de Apoio ao Projeto Integrado: Sistemas Distribuídos
PERÍODO	5
Nº DE CRÉDITOS	15 Teoria, 30 prática e 15 extensão
CARGA HORÁRIA	72
SEMESTRE / ANO	5
PROFESSOR	Carlos Eduardo de Barros Paes

Consultoria Especializada de Apoio ao Projeto Integrado: Sistemas Distribuídos

Ementa

Conceitos básicos de sistemas distribuídos; aplicações distribuídas e paralelas; características básicas da Programação Distribuída; modelo de processos e troca de mensagens; algoritmos distribuídos; conceitos básicos de programação paralela; algoritmos paralelos; classes de linguagens para a Programação Distribuída. MapReduce. Spark. Resilient Distributed Datasets. Desenvolvimento de ação extensionista voltada às comunidades de software de código aberto. Ações extensionistas como resultado do desenvolvimento de projetos de produção de diagnóstico, elaboração de oficinas, apresentação de palestras ou outro tipo de intervenção, a depender do tema e da necessidade do público/ organização atendida. Previsão de produto a ser disponibilizado em repositórios públicos de software como por exemplo o GitHub.

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 à sincronização em sistemas distribuídos
- Aplicar esses conhecimentos no processo de desenvolvimento de sistemas computacionais distribuídos e paralelos usando tecnologias atuais.
- Analisar e avaliar programas desenvolvidos para um ambiente de execução constituído por diversas unidades de processamento operando em paralelo

Conteúdo Programático

Data / Semana	Conteúdo por Aula	Metodologia ou Estratégias de Ensino (Metodologias Ativas, projetos, sala de aula invertida, trabalhos em grupo, entrevistas, seminários)	Recursos Tecnológicos ou Físicos (Plataforma / Software / Aplicativos / Salas de Aula específicas / Laboratórios / Equipamentos)	Observações
Semana 01	Apresentação da disciplina de SD e plano de ensino 2023 Introdução aos Sistemas Distribuídos e Computação Paralela Prática: Introdução ao Linux e Comandos Básicos de Shell	Prática: trabalho em grupo	Sistema Operacional Linux e Linguagem de Programação Python	

Semana 02	Revisão de Sistemas Operacionais Computação paralela e os modelos de máquinas: Classificação de Flynn Introdução a multiprogramação e programação concorrente/paralela (processos e threads) Prática: Experimento de programas concorrentes usando Python	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto Programação Paralela	Sistema Operacional Linux e Linguagem de Programação Python	
Semana 03	Arquiteturas com memória compartilhada e Distribuída Prática: Exemplos de programação sequencia e concorrente/paralela usando programação concorrente em Python	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto Programação Paralela	Sistema Operacional Linux e Linguagem de Programação Python	
Semana 04	Introdução a computação paralela com GPU e CUDA Laboratório: Apresentação de exemplos de programação paralela com CUDA usando Python	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto Programação Paralela	Sistema Operacional Linux e Linguagem de Programação Python	
Semana 05	Arquitetura de Sistemas Distribuídos e Paradigmas de Comunicação em SD Visão Geral do Modelo Cliente/Servidor Tecnologias de comunicação para SD: RPC/RMI, gRPC, SOA, Message-oriented middleware (MOM), Rest/Restful, Sockets, WebSockets) Laboratório: Preparação dos seminários Tecnologias de Comunicação SD	Seminários SD - Tecnologia de comunicação em SDs Apoio ao Projeto Integrador do 5o Período	Sistema Operacional Linux e Linguagem de Programação Python	
Semana 06	Tecnologias de comunicação para SD: RPC/RMI, gRPC, SOA, Message-oriented middleware (MOM), Rest/Restful, Sockets, WebSockets) Laboratório: Preparação dos seminários Tecnologias de Comunicação SD	Seminários SD - Tecnologia de comunicação em SDs Apoio ao Projeto Integrador do 5o Período	Sistema Operacional Linux e Linguagem de Programação Python	
Semana 07	Apresentação dos Seminários: Tecnologias de comunicação para SD: RPC/RMI, gRPC, SOA, Message-oriented middleware (MOM), Rest/Restful, Sockets, WebSockets)	Seminários SD - Tecnologia de comunicação em SDs Apoio ao Projeto Integrador do 5o Período		
Semana 08	Introdução ao Processamento de Big Data Distribuído: Ecossistema Spark e Hadoop Arquitetura e Componentes da Plataforma Smapark Context e Session Formatos de Big Data Prática: Inicialização dos servidores Spark e introdução ao PySpark	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop e Spark	
Semana 09	Dataframes e RDDs Principais Ações e Transformações Prática: Exemplos de processamento de RDDs e Dataframes usando PySpark	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop e Spark	
Semana 10	Spark SQL Outras Fontes de Dados Criando aplicações em Spark Prática: Exemplos de Spark-sql, Joins com Dataframe e SQL	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop e Spark	
Semana 11	Machine Learning com Spark Regressão e Classificação Pipelines	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop e Spark	

Semana 12	Spark Structured Streaming Visão Geral de Streaming Casos Práticos	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop e Spark	
Semana 13	Aspectos Gerais de Otimização Particionamento e Buckering no Spark Cache e Persistência no Spark	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop e Spark	
Semana 14	Clusterização em Sparck Parática: Clone das máquinas virtuais, definição dos nós de processamento (master e workers), configuração do clustes.	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop e Spark	
Semana 15	Clusterização em Sparck Prática: Clone das máquinas virtuais, definição dos nós de processamento (master e workers), configuração do clustes.	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop e Spark	
Semana 16	Hadoop Overview e História Haddop Cores: HDFS e MapReduce Prática: Inicialização do Hadoop, configuração do HDFS e comandos HDFS	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop e Spark	
Semana 17	Programando Hadoop com Python e Spark Introdução ao Ambari Prática: Exemplos de programas usando Python	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop e Spark	
Semana 18	Usando BD Relacional e Não Relacional com Hadoop Introdução ao Hive Prática: Integração do Hadoop com os BDs MySQL e MongoDB	Prática: trabalho em grupo Aprendizagem baseada em Projetos: Projeto integrador do 5o período	Sistema Operacional Linux, Linguagem de Programação Python, Hadoop, Spark, MongoDB	

Avaliação				
Data da Avaliação	Forma de Avaliação (Oral / Escrita / Seminário / Projeto / Entrega de Relatório / outro (indicar))	Tipo: Individual / Grupo	Pesos (caso houver)	Recurso tecnológico (quando necessário) Plataforma/ Softwares/Aplicativos, etc)
Apresentação dos projetos finais	Projeto Integrador	Individual e Grupo		

Bibliografias Básica e Complementar
Básica: Como Consta no PPC
CZECH, J. Z. Introduction to Parallel Computing. Cambridge: Cambridge University Press, 2017
TANENBAUM, A. S.; VAN STEEN, M. Distributed systems: principles and paradigms. 2. ed

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

Complementar: Indicadas pelo professor para Consulta

DAMJI, J.; WENIG, B.; DAS, T.; LEE, D. Learning Spark: Lightning-Fast Data Analytics. 2. ed. EUA: O'Reilly Media, 2020.

CHAMBERS, B.; ZAHARIA, M. Spark: The Definitive Guide: Big Data Processing Made Simple. EUA: O'Reilly Media, 2018.

TANDOM, A. et al. Advanced Analytics with Pyspark: Patterns for Learning from Data at Scale Using Python and Spark. EUA: O'Reilly Media, 2022.

PARSIAN, M. Data Algorithms with Spark: Recipes and Design Patterns for Scaling Up Using Pyspark. EUA: O'Reilly Media, 2022.

LEE, D.; DRABAS, T. PySpark Cookbook: Over 60 recipes for implementing big data processing and analytics using Apache Spark and Python. 1 ed. [S.l.]: Packt Publishing, 2018.