

Programa Analítico de Disciplina

CDP 200 - Big Data I

Campus Rio Paranaíba -

Catálogo: 2026

Número de créditos: 4

Carga horária semestral: 60h

Carga horária semanal teórica: 4h

Carga horária semanal prática: 0h

Carga horária de extensão: 0h

Semestres: I e II

Objetivos

- Compreender os fundamentos de Big Data e seus desafios
- Dominar o ecossistema Hadoop e suas componentes
- Implementar algoritmos usando o modelo MapReduce
- Desenvolver aplicações com Apache Spark
- Conhecer e aplicar bancos de dados NoSQL
- Projetar pipelines de ingestão e processamento de dados
- Desenvolver habilidades para processamento distribuído

Ementa

Fundamentos de Big Data. Arquiteturas e ecossistema Hadoop. Sistema de arquivos HDFS. Modelo de programação MapReduce. Processamento distribuído com Apache Spark. Spark RDD e DataFrames. Otimização de consultas em grandes volumes de dados. Conceitos de data lakes e data warehouses. NoSQL: bancos de dados chave-valor, documentais, colunares e de grafos. Técnicas de ingestão e processamento de dados em streaming. Introdução a ferramentas de orquestração de pipelines.

Pré e correquisitos

SIN 216

Oferecimentos obrigatórios

Não definidos

Oferecimentos optativos

Curso	Grupo de optativas
Ciência de Dados e Inteligência Artificial	Grupo Geral
Sistemas de Informação	Geral

A autenticidade deste documento pode ser conferida no site <https://siadoc.ufv.br/validar-documento> com o código: 4FE2.7MPW.GXSJ

CDP 200 - Big Data I

Conteúdo					
Unidade	T	P	ED	Pj	To
1. Introdução a Big Data 1. Conceitos fundamentais: Volume, Velocidade, Variedade, Veracidade 2. Arquiteturas de Big Data 3. Casos de uso e aplicações 4. Desafios tecnológicos 5. Ecosistema Hadoop 6. Computação distribuída vs paralela	8h	0h	0h	0h	8h
2. Hadoop e HDFS 1. Arquitetura Hadoop 2. Hadoop Distributed File System (HDFS) 3. NameNode e DataNodes 4. Replicação e tolerância a falhas 5. Comandos HDFS 6. YARN e gerenciamento de recursos 7. Monitoramento do cluster	10h	0h	0h	0h	10h
3. Modelo MapReduce 1. Fundamentos do MapReduce 2. Mappers e Reducers 3. Shuffle e Sort 4. Otimização de jobs MapReduce 5. Implementação de algoritmos clássicos 6. MapReduce vs alternativas modernas	10h	0h	0h	0h	10h
4. Apache Spark 1. Arquitetura do Spark 2. Resilient Distributed Datasets (RDDs) 3. Transformações e ações 4. Spark DataFrames e Datasets 5. Spark SQL 6. Otimização com Catalyst Optimizer	12h	0h	0h	0h	12h
5. Bancos de Dados NoSQL 1. Teorema CAP 2. Modelos de dados: chave-valor, documento, colunar, grafos 3. Cassandra e HBase	10h	0h	0h	0h	10h

A autenticidade deste documento pode ser conferida no site <https://siadoc.ufv.br/validar-documento> com o código: 4FE2.7MPW.GXSJ

4.MongoDB 5.Redis 6.Neo4j 7.Casos de aplicação					
6. Ingestão e Processamento Básico	10h	0h	0h	0h	10h
1.Ferramentas de ingestão: Sqoop, Flume 2.Conceitos de Data Lakes 3.Introdução ao processamento streaming 4.Projeto prático integrador 5.Apresentação de resultados					
Total	60h	0h	0h	0h	60h

Teórica (T); Prática (P); Estudo Dirigido (ED); Projeto (Pj); Total (To);

Planejamento pedagógico	
Carga horária	Itens
Teórica	Aulas expositivo dialogadas
Prática	Seminários e Aulas em laboratório de informática
Estudo Dirigido	<i>Não definidos</i>
Projeto	<i>Não definidos</i>
Recursos auxiliares	<i>Não definidos</i>

CDP 200 - Big Data I

Bibliografias básicas

Descrição	Exemplares
WHITE, Tom. Hadoop: The Definitive Guide. 4. ed. Sebastopol: O'Reilly Media, 2015.	0
AMARAL, Fernando. Introdução à Ciência de Dados. 1. ed. Rio de Janeiro: Alta Books, 2016. 320 p. ISBN 978-8576089346.	0
ACHARYA, S., CHELLAPPAN, S., "Big Data Analytics" Wiley 2015.	0

Bibliografias complementares

Descrição	Exemplares
HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Data Mining: Concepts and Techniques. 3. ed. Waltham: Morgan Kaufmann, 2012.	0
SADALAGE, Pramod J.; FOWLER, Martin. NoSQL Essencial: Um Guia Conciso para o Mundo Emergente da Persistência Poliglota. São Paulo: Novatec Editora, 2013. 216 p. ISBN 978-85-7522-338-3.	0
DAMJI, Jules S.; WENIG, Brooke; DAS, Tathagata; LEE, Denny. Learning Spark: Lightning-Fast Data Analytics. 2. ed. Sebastopol: O'Reilly Media, 2020. 397 p. ISBN 978-1492050049.	0
INMON, W. H. Building the Data Warehouse. 4. ed. Hoboken: Wiley, 2005.	0
NETTO, Amílcar; MACIEL, Francisco. Python para Data Science: e Machine Learning Descomplicado. 1. ed. Rio de Janeiro: Alta Books, 2021.	0