

# Engenharia de Dados

*Apresentação da disciplina*

DCA0132 - Engenharia de Dados

Prof. Carlos M. D. Viegas

**DCA**

Departamento de Engenharia de Computação e Automação  
Universidade Federal do Rio Grande do Norte

**UFRN**  
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

# Apresentação da disciplina

- Engenharia de Dados (DCA0132)
  - 60 horas
  - Horário 24M34
- Metodologia
  - Aulas expositivas
  - Atividades práticas
- Formas de avaliação
  - Provas teóricas
  - Práticas de laboratório
  - Trabalhos de implementação

# Apresentação da disciplina

- Requisitos básicos “exigidos”
  - Habilidade para programar em linguagem *Python*
    - Tutorial: <https://wiki.python.org.br/AprendaProgramar>
  - Realizar o curso *Big Data Fundamentals* disponível em: <https://www.datascienceacademy.com.br/course/big-data-fundamentos-3>
  - Estudar e compreender as competências de um Engenheiro de Dados: <https://awari.com.br/guia-engenharia-de-dados/>
  - Conhecimentos intermediários de sistemas operacionais
    - GNU/Linux e MS Windows
  - Ser curioso e proativo

# Caracterização da disciplina

- Ementa:

- Estudo das estruturas de dados para **manipulação de dados em grande escala**, mineração de dados, **bancos de dados não relacionais**, ferramentas para gerenciamento de **dados distribuídos**, computação em nuvem e **virtualização**.

- Objetivos:

- Apresentar os conceitos sobre o **futuro das aplicações baseadas em dados**, combinando conhecimentos de sistemas de bancos de dados, computação distribuída, processamento de dados, criação de pipelines de dados, virtualização de processos, escalonamento e monitorização de tarefas.

- Competências e Habilidades:

- O aluno terá como habilidades principais a capacidade para **planejar e preparar a infraestrutura e arquitetura de dados** de uma organização, podendo projetar, construir, integrar e manter banco de dados relacionais e não relacionais, dados não estruturados e dados coletados de diversas fontes, bem como conhecer as técnicas e ferramentas para processamento de dados em larga escala.

# Caracterização da disciplina



# Conteúdo programado

## Unidade I

- Introdução à Engenharia de Dados
- Virtualização de processos
  - Hypervisores e Containers
- Introdução ao ecossistema Apache Hadoop
  - Criação e gerenciamento de clusters de dados
  - Processamento de dados com MapReduce e Apache Spark

# Conteúdo programado

## Unidade II

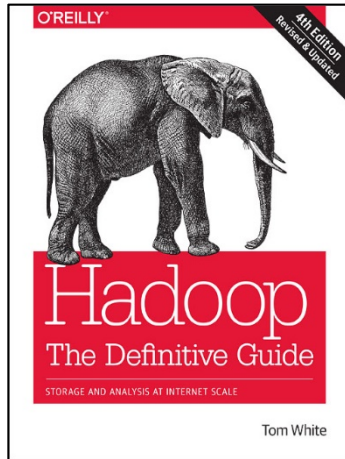
- Bancos de dados não relacionais
  - Introdução ao NoSQL
  - Instalação e configuração de bancos de dados NoSQL

## Unidade III

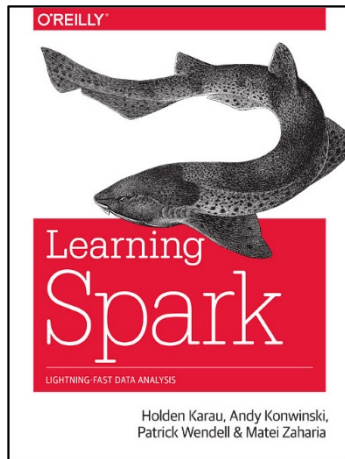
- Gerenciamento de fluxos de trabalho para pipelines de dados
  - Criação de pipelines de dados
  - Escalonamento e monitorização de tarefas

# Bibliografia

- Principal



- WHITE, Tom. Hadoop: The Definitive Guide. O'Reilly, 2015. 756 p.



- KARAU, Holden; KONWINSKI, Andy; WENDELL, Patrick; ZAHARIA, Matei. Learning Spark: lightning-fast data analysis. O'Reilly, 2015. 276 p.

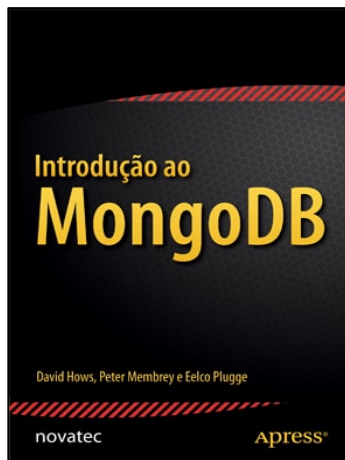


# Bibliografia

- Complementar



- BENGFORT, Benjamin; KIM, Jenny. Analítica de Dados com Hadoop: uma introdução para cientistas de dados. Novatec, 2016. 352 p.



- HOWS, David; MEMBREY, Peter; PLUGGE, Eelco. Introdução ao MongoDB. Novatec, 2015. 168 p.

# Links úteis

- Cognitive Class
  - <https://cognitiveclass.ai/courses>
- Dataquest
  - <https://www.dataquest.io/data-science-courses/>
- DataCamp
  - <https://www.datacamp.com/courses/all>
- Data Science Academy
  - <https://www.datascienceacademy.com.br/pages/todos-os-cursos-dsa>

# Atendimento e dúvidas

- Grupo no Telegram
  - <https://tinyurl.com/engdados20222>
- Outros assuntos
  - [viegas@dca.ufrn.br](mailto:viegas@dca.ufrn.br)
- Página do Professor (DCA)
  - Acessar <https://www.dca.ufrn.br/~viegas>  
Ensino > DCA0132 - Engenharia de Dados