



# GUIA DATA MASTERS

ENGENHEIRO DE DADOS



# DATA MASTERS

## ENGENHEIRO DE DADOS

Olá, seja muito bem-vindo ao Guia **DATA MASTERS** de Engenheiro de Dados! Aqui, você encontrará uma seleção de conteúdos recheados para auxiliá-lo na sua certificação.

O Engenheiro de Dados tem como principal função preparar a infraestrutura de dados, desde os repositórios de dados até suas integrações e distribuição, além de otimizar a performance de queries complexas sobre os dados do Banco e assegurar que eles estão facilmente acessíveis.

Preparamos para você artigos, livros, cursos, vídeos e palestras que vão lhe proporcionar um aprendizado realmente incrível.

## VOCÊ ENCONTRARÁ...

### **06** Sistemas Operacionais

- Conceitos e prática da arquitetura de sistemas operacionais.

### **08** Gestão de Eventos

- Patterns de mensageria;
- Processamento de eventos.

### **10** Programação

- Lógica;
- Boas práticas de programação;
- Bibliotecas e APIs de Machine Learning.

### **13** Manipulação de Dados

- Conceitos de modelagem de dados;
- Comandos SQL e NoSQL;
- Conceitos de Big Data.

### **16** Big Data

- Instalação e uso de ferramentas de armazenamento;
- Tableau;
- Watson Analytics;
- Sisense;
- Oracle Bi;
- Movimentação e processamento de dados .

### **20** Ingestão de Dados

- Conceitos e prática para saneamento de dados;
- Data quality;
- ETL.

### **22** Cursos e Livros

- K-means;
- Algoritmos hierárquicos aglomerativos e divisivos;
- Latent Dirichlet Allocation.

# SUMÁRIO

---

## ENGENHEIRO DE DADOS

**06** Sistemas Operacionais

**08** Gestão de Eventos

**10** Programação

**13** Manipulação de Dados

**16** Big Data

**20** Ingestão de Dados

**22** Cursos e Livros

# ICONOGRAFIA

---

No conteúdo deste e-book, você encontrará diversos materiais, dentre eles opções pagas e abertas.

Abaixo você confere os significados de cada ícone e ao visualizá-los, basta clicar para acessar.



Artigo



PDF



Curso



Vídeo



Livro



Nível Básico



Nível Intermediário



Nível Avançado

# SISTEMAS OPERACIONAIS

Nessa seção do Guia **DATA MASTERS**, você encontrará uma série de conteúdos que auxiliarão na conceituação e na prática da arquitetura de sistemas operacionais.

## CONCEITOS E PRÁTICA DA ARQUITETURA DE SISTEMAS OPERACIONAIS



● ○ ○

How Do Operating Systems Work?

🕒 4min 



● ○ ○

Organização e arquitetura de computadores

🕒 2h



● ○ ○

Sistemas Operacionais

🕒 8min



● ○ ○

Sistemas Operacionais  
- UFSM

🕒 1h30



● ○ ○

What is an operating  
system

🕒 20min 🇬🇧



● ○ ○

Structures of Operating  
System

🕒 17min 🇬🇧



● ○ ○

Sistemas Operacionais  
Modernos Editora Pearson

Pago R\$ 223,30



● ○ ○

Guia Foca GNU/Linux



# GESTÃO DE EVENTOS

---

Dentro desse setor serão apresentados patterns de mensageria e as etapas que compõem o processamento de eventos.

## PATTERNS DE MENSAGERIA



Mensageria de seus sistemas

🕒 25min



Enterprise Integration Patterns  
Editora Addison Wesley

Pago R\$ 129,36



## PROCESSAMENTO DE EVENTOS




Eventos: definição e processamento

🕒 40min



What is event processing?

🕒 10min 



Complex Event Processing

🕒 1h30

# PROGRAMAÇÃO

Agora, vamos lhe apresentar os seguintes conteúdos: Conceitos Básicos de Lógica; Algoritmos; Linguagem de Programação; Boas Práticas e algumas bibliotecas e APIs de Machine Learning.

Fique tranquilo, pois, se você não tiver familiaridade com a linguagem de programação, nessa seção você encontrará tudo o que precisa saber sobre Python, e como está presente no cotidiano de um Engenheiro de Dados.

## LÓGICA



● ○ ○

Introdução à programação  
com Python Editora Novatec

Pago R\$ 73,00



● ○ ○

Learning Python  
Editora O'Reilly

Pago R\$ 244,95



● ○ ○

Lógica de Programação  
e Algoritmos

🕒 3h

Pago R\$ 270,01



● ○ ○

Lógica de programação

🕒 1h30min

## BOAS PRÁTICAS DE PROGRAMAÇÃO



Código limpo com Clean Code

🕒 20min



Design Patterns com Python

🕒 1h



Design Patterns in Python by Peter Ullrich

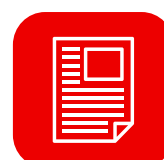
🕒 28min

## BIBLIOTECAS E APIS DE MACHINE LEARNING



TOP 20 bibliotecas Python para Data Science

🕒 30min



Melhores bibliotecas de machine e deep learning

🕒 30min




11 ferramentas open source para machine learning

🕒 15min



Machine learning APIs

🕒 2h30min 



Scikit-Learn Machine  
learning APIs



Tensorflow Machine  
learning APIs

# MANIPULAÇÃO DE DADOS

Em Manipulação de Dados vamos analisar requisitos, definir o modelo conceitual e modelo lógico. Por fim, entender como aplicar o modelo físico. Ficou confuso? Não se preocupe! Nessa seção vamos abordar os conceitos de NoSQL e modelagem de dados e como aplicar os comandos da Linguagem SQL.

## CONCEITOS DE MODELAGEM DE DADOS



● ○ ○

Entenda o que é a modelagem de banco de dados

🕒 15min



● ○ ○

Modelagem de Dados -  
Conceitos de Bancos de Dados

🕒 9h



● ○ ○

Introdução a Sistemas de Bancos de Dados Editora Elsevier

Pago R\$ 408,00



● ○ ○

Sistemas de Banco de Dados  
Editora Pearson

Pago R\$ 263,84

## COMANDOS SQL E NoSQL



● ○ ○

Comandos básicos em SQL insert  
update delete e select

🕒 20min



● ○ ○

Guia completo de SQL

🕒 2h



● ○ ○

Introdução a SQL: Consulta e  
gerenciamento de dados

🕒 1h30



● ○ ○

SQL Tutorial

🕒 2h 



● ○ ○

O que é NoSQL

🕒 20min



● ○ ○

Top 6 NoSQL data bases

🕒 20min

## CONCEITOS DE BIG DATA



● ○ ○

Big Data

🕒 6min



● ○ ○

Big data - O que é e qual sua importância?

🕒 30min



● ○ ○

O que é Big Data - Conceitos básicos

🕒 18min



# BIG DATA

---

Você sabe o que é Big Data? Este é um termo que trata de grandes conjuntos de dados que precisam ser processados e armazenados.

Neste momento, nos deteremos no processo de instalação e uso de ferramentas de armazenamento, além da movimentação e do processamento de dados.

## INSTALAÇÃO E USO DE FERRAMENTAS DE ARMAZENAMENTO



Análise de Regressão



Chef Foundations -  
Linux & Windows



## TABLEAU



● ● ● ○

Complete tutorial TABLEAU

🕒 9h 

## HADOOP



● ● ● ○

Hadoop: The Definitive Guide  
Editora O'Reilly

Pago R\$ 158,55 



● ● ● ○

Hadoop Application Architectures:  
Designing Real-World Big Data  
Applications Editora O'Reilly

Pago R\$ 74,89 



● ● ● ○

Plataforma Hadoop e framework  
de aplicação

🕒 20h



● ● ● ○

Big Data Essentials: HDFS,  
MapReduce and Spark RDD

🕒 44h



● ● ● ○

Modern Big Data Processing  
with Hadoop Packt Publishing





● ● ● ○

Business Intelligence

 Pago R\$ 579,99

## MOVIMENTAÇÃO E PROCESSAMENTO DE DADOS



Kafka Hadoop Integration

🕒 20min 



Talend and Apache Spark:  
A Technical Primer

🕒 30min 



Kafka Tutorial | Apache Kafka  
Tutorial For Beginners

🕒 17min 



Etapas de Big Data

🕒 1h




Integrating Hadoop with  
SQL Server

🕒 40min 



Data processing

🕒 10min 




Impala





5 different types of data processing

🕒 50min 



## Introduction to Data and Data Processing

🕒 12h



# INGESTÃO DE DADOS

Por fim, em Ingestão de Dados, iremos destrinchar os conceitos e práticas para Saneamento de dados, Data Quality e ETL (Extração, Transformação, Carregamento).

## CONCEITOS E PRÁTICA PARA SANEAMENTO DE DADOS



Data sanitization terminology

🕒 1h30 🇬🇧



Data sanitization definition

🕒 15min 🇬🇧



Data sanitization

🕒 40min 🇬🇧



Data sanitization methods

🕒 30min 🇬🇧

## DATA QUALITY



Data Quality: O que é, quais benefícios e como adotar a prática?

🕒 20min



Data Quality: Como está a qualidade dos nossos dados?

🕒 40min



Data Quality

🕒 20min



Data Quality

🕒 30min 

## ETL



What is ETL?

🕒 3min 



O que é ETL?

🕒 25min



Entendendo o processo de ETL

🕒 15min

# CURSOS E LIVROS

Para você que aprecia um estudo de forma linear, como uma trilha de aprendizado, separamos alguns Cursos e Livros para complementar no seu processo de aprendizagem.

Os cursos abordam de forma transversal os assuntos do Guia **DATA MASTERS**, mas indicamos que você também estude todos os conteúdos, pois assim você vai garantir a total compreensão dos temas.

Bons estudos!



## Data Science de A a Z - Da extração a exibição dos Dados

🕒 32h30 **Pago R\$ 579,99**

- Utilizar estatística aplicada;
- Análise de Dados com Python;
- Definir o que é Machine Learning, Big Data e Data Science;
- Gráficos com Python.



## Formação cientista de dados

🕒 376h **Pago R\$ 3515,00**

- Big Data Analytics com R e Microsoft Azure Machine Learning, Versão 2.0;
- Big Data Real-Time Analytics com Python e Spark;
- Business Analytics;
- Engenharia de Dados com Hadoop e Spark;
- Machine Learning;
- Visualização de Dados e Design de Dashboards;
- Preparação para Carreira de Cientista de Dados.





● ● ○

## Bootcamp Cientista de Dados: 10 aplicações Web Completas

🕒 12h Pago R\$ 309,99

- Implemente aplicação de Machine Learning;
- Desenvolva soluções para problemas de classificação;
- Use distribuições de probabilidade para prever acontecimentos;
- Crie Dashboards de Análise de Dados Exploratória;
- Aplique Algoritmos de Inteligência Artificial;
- Avalie modelos de previsão de Séries Temporais.



● ● ○

## The Data Science Course 2019: Complete Data Science Bootcamp

🕒 12h Pago R\$ 579,99



● ● ○

## Data Science Harvard

🕒 38h 

- Fundamental R programming skills;
- Statistical concepts such as probability, inference, and modeling and how to apply them in practice;
- Gain experience with the tidyverse, including data visualization with ggplot2 and data wrangling with dplyr;
- Become familiar with essential tools for practicing data scientists such as Unix/Linux, git and GitHub, and RStudio;
- Implement machine learning algorithms;
- In-depth knowledge of fundamental data science concepts through motivating real-world case studies;
- Program Overview.



## Programa de cursos integrados Big Data

🕒 7h

- Introdução a Big Data;
- Big Data Modeling and Management Systems;
- Big Data Integration and Processing;
- Aprendizado de máquina com Big Data.



## My Mooc

Pesquisa pela competência que quer desenvolver



## Data Science Para Negócios

Editora Alta Books

Pago R\$ 87,90



## Data Science do zero

Editora Elsevier/Alta Books

Pago R\$ 76,90

