

Módulos 1 e 2

## **Engenharia de Dados**



- O que é um engenheiro de dados?
- Como criar uma caixa de ferramentas para ingestão e processamento de dados para nos atender no dia a dia independente da nossa profissão.

**Módulo 1:** Fundamentos de Engenharia de Dados

**Módulo 2:** Fundamentos de Ingestão de Dados

---

Módulos 3 a 6

## **Fundamentos de ingestão de dados**



- Fundamentos, começando com processos para ingestão de dados de APIs e também com crawlers.
- Boas práticas: retentativas, checkpointing e logs. Neste módulo faremos também uma revisão de conceitos python e de SQL.

**Módulo 3:** SQL

**Módulo 4:** Capturando dados de uma API

**Módulo 5:** Capturando dados com crawlers

**Módulo 6:** Testes e Jenkins

---

Módulos 7 a 9

## **Testes e Ambientes de Trabalho**



- Construir ambientes de trabalho – seja localmente ou em uma pequena estrutura – com ferramentas como Jenkins e Docker.
- Iniciar abordagem em um ambiente AWS com a criação de conta na Amazon e o deploy de uma função lambda.

**Módulo 7:** Introdução à AWS

**Módulo 8:** AWS Lambda

**Módulo 9:** Git/Github

---

Módulos 10 e 11

## **CI/CD Developer tools**



- Ferramentas para desenvolvimento de software que nos ajudam a gerenciar versões

de código e a automatizar testes e deploys.

- Entender o git flow e aprender boas práticas para desenvolvimento de código em equipes.
- Integração contínua com o GitHub Actions.
- Executar testes de forma automatizada.
- Deploy de recursos automaticamente.

#### **Módulo 10: CI/CD**

#### **Módulo 11: Ingestão de Dados**

---

Módulos 12 a 20

### **Data Lake na AWS**



- Conceitos e boas práticas de data lakes.
- Criação de lake na AWS utilizando ferramentas como S3, Glue, Athena e Redshift Spectrum.
- Estudar processamento de dados utilizando Spark.
- Aprender a orquestrar pipelines de dados utilizando Apache Airflow.

#### **Módulo 12: Data Lakes**

#### **Módulo 13: AWS Glue + AWS Athena**

#### **Módulo 14: Mensageria**

#### **Módulo 15: Redshift e Spectrum**

#### **Módulo 16: Spark no Databricks**

#### **Módulo 17: Spark Streaming no Databricks**

#### **Módulo 18: Airflow**

#### **Módulo 19: Terraform e Cloudformation**

#### **Módulo 20: Data Engineering Portfolio**

---

Módulo 21

### **Cloudformation e Terraform**



- Aprender duas ferramentas para gerenciamento de Infraestrutura como código, essenciais para a criação de projetos de engenharia de dados escaláveis e fáceis de manter.
- Como criar recursos na AWS através de código utilizando Cloudformation e Terraform.

---

Módulo 22

### **Live sobre carreira**



Live com convidados especiais contando sobre sua carreira e dando dicas de desenvolvimento de portfólio, como se candidatar e atuar em entrevistas, projetos e oportunidades no dia a dia.

---