

2026

---

FORMAÇÃO

ENGENHARIA DE DADOS E  
DATA LAKE ENTERPRISE



Not a   
— lab.  
*Nós fazemos!*

---

NOTALAB.COM.BR  
BRASIL

## ENGENHARIA DE DADOS E BIG DATA LAKE ENTERPRISE COM DATABRICKS, AWS, GCP E AZURE

Vá além do Data Lake, construindo um oceano de dados. Impulsione a sua carreira



14 cursos em 1



Videos e Materiais



Databricks, AWS,  
GCP e Azure



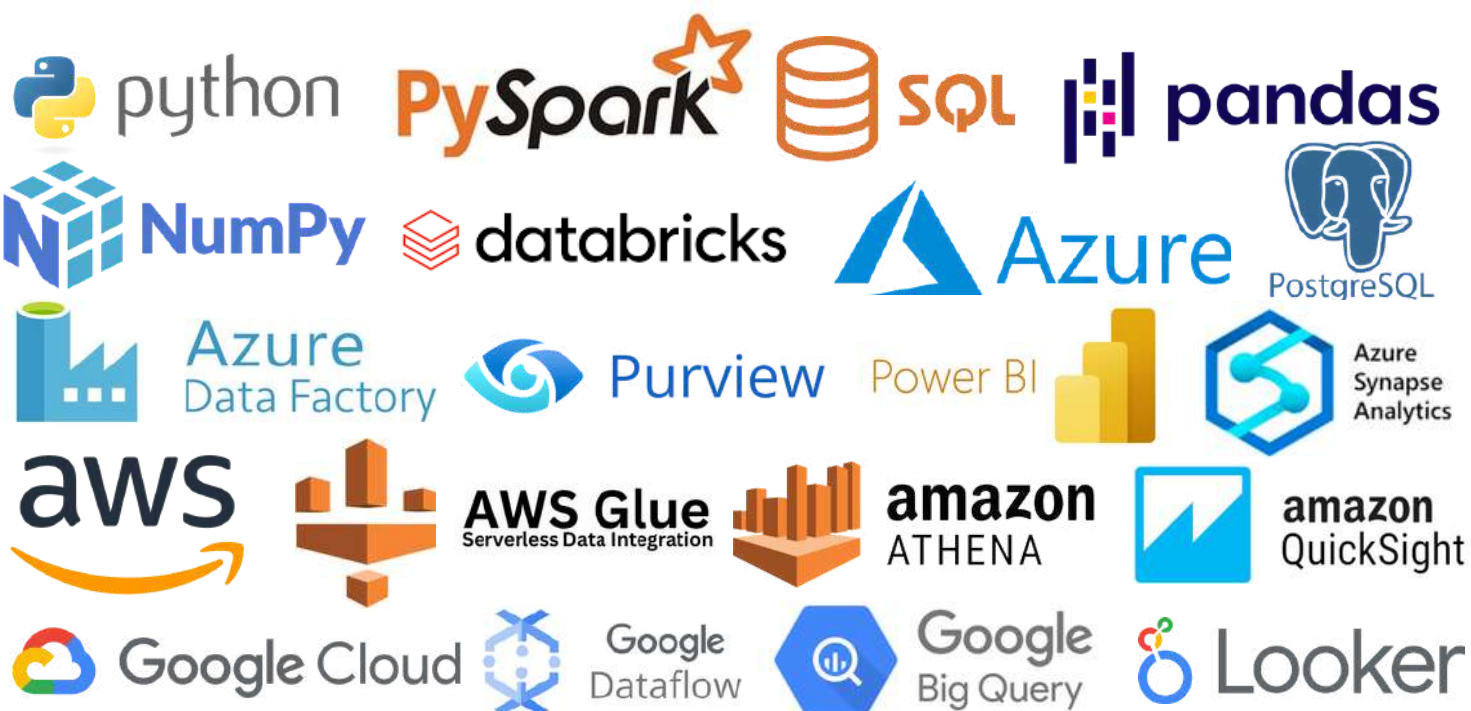
Impulsione sua  
carreira

Aprenda as principais técnicas e linguagens da Engenharia de Dados (Python, SQL e PySpark). Entenda as diferenças entre a estruturação de dados tradicional até os conceitos de Data Warehouse e Big Data. Saiba como construir um Data Lake de Sucesso e ir além com o novo conceito de Oceano de Dados!

### Público Alvo:

Estudantes de Tecnologia, Desenvolvedores(as) e Programadores(as), Cientistas de Dados, Arquitetos(as) de Dados, público de tecnologia em geral que tem vontade de entender o mundo do Big Data.

### TECNOLOGIAS QUE VOCÊ VAI APRENDER



## PROFESSOR

---



**Rodolfo Moreira**

AI & Data Manager | Artificial Intelligence  
Professor | AI Manager | Data Science Mana...



✉ [contato@rodolfomoreira.com.br](mailto:contato@rodolfomoreira.com.br)

📞 11 9 1600-4334

🌐 [in/rodolfofordriguesmoreira/](https://in.linkedin.com/in/rodolfofordriguesmoreira/)

Com 16 anos de experiência em tecnologia, atuo como especialista em Dados e Inteligência Artificial, com foco na construção de aplicações de dados ponta a ponta — da arquitetura à entrega de valor.

Graduado em Análise e Desenvolvimento de Sistemas e com MBA em Inteligência Artificial & Machine Learning, também leciono disciplinas de Dados e IA, conectando teoria e prática.

Minha missão é transformar dados em produtos inteligentes, capazes de gerar eficiência e novas oportunidades de negócio.

Entre os principais objetivos dos projetos estão:

- Redução de custos e aumento de eficiência;
- Automação de rotinas e processos;
- Melhoria no atendimento e experiência do cliente;
- Aumento de conversão e receita.

## PROJETOS DE DADOS REALIZADOS EM GRANDES EMPRESAS DO BRASIL



## CURSOS DA FORMAÇÃO



### INTRODUÇÃO DATA LAKE ENTERPRISE

Este curso apresenta os fundamentos essenciais para entender como empresas modernas estruturam, organizam e extraem valor de seus dados em escala. Você aprenderá como funciona a maturidade analítica dentro das organizações, como o conceito de “oceano de dados” torna empresas mais competitivas, quais são os principais perfis que compõem equipes de dados e como essas equipes atuam na prática dentro do mercado.

#### AULAS

##### **Aula 1 – Maturidade de dados dentro das empresas**

Entenda o caminho que as organizações percorrem até se tornarem data-driven e como evoluem seus processos, ferramentas e cultura analítica.

##### **Aula 2 – Como uma empresa pode ser competitiva utilizando o conceito de oceano de dados?**

Descubra como dados centralizados, acessíveis e integrados impulsionam decisões estratégicas, inovação e vantagem competitiva.

##### **Aula 3 – Perfis de dados dentro das equipes**

Conheça as principais funções do ecossistema de dados – Engenharia, Analytics, Ciência, Arquitetura, Governança – e seus papéis no dia a dia.

##### **Aula 4 – Atuação das equipes e a realidade nas empresas**

Veja como as equipes se organizam, colaboram e operam na prática, entendendo processos reais, desafios e responsabilidades de cada área.



## PROGRAMAÇÃO PARA DADOS COM PYTHON

Este curso apresenta os fundamentos essenciais de Python aplicados ao universo de dados. Você aprenderá desde os conceitos básicos da linguagem até estruturas mais avançadas utilizadas no dia a dia de profissionais de Engenharia de Dados, Ciência de Dados, Automação e IA. Ao final, você estará apto a escrever seus próprios scripts, manipular dados com bibliotecas essenciais e construir a base necessária para avançar para módulos mais técnicos da formação.

### AULAS

#### **Aula 1 – Introdução ao Python**

Entenda o que é Python, por que ele é tão utilizado em dados e como funciona sua sintaxe básica.

#### **Aula 2 – Jupyter Notebook no Google Colab e primeiras bibliotecas**

Aprenda a usar notebooks na nuvem e faça seus primeiros testes com bibliotecas para análise de dados.

#### **Aula 3 – Bibliotecas Essenciais para Dados**

Conheça as principais ferramentas do ecossistema Python: Pandas, NumPy, Matplotlib, entre outras.

#### **Aula 4 – Python com Interpretador e Visual Studio Code**

Configure seu ambiente local e use o VS Code como IDE profissional.

#### **Aula 5 – Variáveis, Tipos de Dados e Funções**

Explore a base da lógica de programação: tipos, criação de funções e boas práticas.

#### **Aula 6 – Casting e Trabalhando com Datas**

Aprenda conversões de tipos e manipulação de datas, algo fundamental para pipelines e análises.

#### **Aula 7 – Operadores de Comparação**

Domine operadores essenciais para tomadas de decisão em scripts e automações.

#### **Aula 8 – Controladores de Fluxo (If/Else)**

Entenda como controlar o fluxo do código e implementar lógicas condicionais.

#### **Aula 9 – Operadores Lógicos e Laços de Repetição**

Trabalhe com loops, condições combinadas e estruturas repetitivas.

#### **Aula 10 – Expressões Ternárias e Módulos**

Aprenda a escrever código mais limpo com ternários e como organizar o projeto utilizando módulos.



### TRABALHANDO COM CLOUD NA MICROSOFT AZURE

Este curso apresenta os conceitos essenciais para quem está iniciando no mundo da computação em nuvem, com foco prático na Microsoft Azure. Aqui você aprenderá desde os fundamentos de Cloud Computing até a criação, monitoramento e gestão de recursos reais dentro da plataforma.

Ao final, você estará pronto para navegar com segurança na Azure, entender seus serviços principais e começar a construir soluções voltadas para dados.

#### AULAS

##### **Aula 1 – O que é Cloud e Plataformas Azure, AWS e GCP**

Entenda os conceitos fundamentais da nuvem e como as principais plataformas se comparam no mercado.

##### **Aula 2 – IaaS, PaaS e SaaS**

Aprenda a diferença entre os modelos de serviço e quando utilizar cada um dentro de uma arquitetura moderna.

##### **Aula 3 – Criando um Recurso PaaS na Azure**

Veja na prática como criar serviços prontos para uso e focados em produtividade.

##### **Aula 4 – Regiões dos Servidores**

Descubra como funcionam regiões, zonas e distribuição física da Azure pelo mundo.

##### **Aula 5 – Criando um Recurso IaaS na Azure**

Aprenda a criar máquinas virtuais e infraestrutura personalizável para seus projetos.

##### **Aula 6 – Monitorando Custos e Despesas**

Entenda como acompanhar gastos, configurar alertas e evitar surpresas na conta.

##### **Aula 7 – Excluindo um Grupo de Recursos**

Aprenda a encerrar ambientes de forma segura e organizada, liberando custos e mantendo o controle do seu ambiente.



## INTRODUÇÃO BANCO DE DADOS

Este curso apresenta os fundamentos essenciais para quem está iniciando no mundo dos bancos de dados, entendendo desde os conceitos básicos até as primeiras operações práticas em SQL utilizando PostgreSQL.

Ao final, você compreenderá a diferença entre bancos relacionais e não relacionais, instalará seu próprio ambiente de banco de dados e executará operações CRUD – base de qualquer aplicação e pipeline de dados.

### AULAS

#### **Aula 1 – Banco de Dados, SGBDs, SQL vs NoSQL**

Entenda o que são bancos de dados, como funcionam os Sistemas Gerenciadores (SGBDs) e quando utilizar SQL ou NoSQL em projetos de dados.

#### **Aula 2 – Instalando o PostgreSQL**

Aprenda a configurar o PostgreSQL no seu computador e criar o ambiente ideal para testes e desenvolvimento.

#### **Aula 3 – Criando Database e Schema no PostgreSQL**

Veja como organizar seus dados criando bancos, schemas e estruturas iniciais de forma profissional.

#### **Aula 4 – CRUD com PostgreSQL (Criar, Ler, Atualizar e Deletar)**

Execute suas primeiras operações SQL e pratique as ações fundamentais utilizadas em qualquer sistema baseado em dados.



### INTRODUÇÃO BIG DATA

Este curso apresenta uma visão completa e introdutória sobre o universo do Big Data, mostrando como as empresas evoluíram na forma de coletar, armazenar e analisar grandes volumes de dados. Você entenderá os fundamentos dessa área, seus principais desafios e as tecnologias que possibilitam trabalhar com dados em escala.

#### AULAS

##### **Aula 1 – Introdução a Big Data: Abordagem histórica do EIS, BI, DW até o Data Lake**

Conheça a evolução dos sistemas de informação, desde o Enterprise Information System até o Data Lake moderno, entendendo como chegamos ao cenário atual do Big Data.

##### **Aula 2 – Os 5 Vs do Big Data**

Explore os pilares que definem Big Data: Volume, Velocidade, Variedade, Veracidade e Valor – e como cada um impacta o dia a dia dos projetos.

##### **Aula 3 – Hadoop vs Spark**

Entenda as diferenças entre os dois ecossistemas mais importantes do processamento distribuído, suas vantagens, limitações e quando usar cada um.



### DATABRICKS E ENGENHARIA DE DADOS

Este curso apresenta uma visão moderna e prática sobre o Databricks – a principal plataforma Lakehouse do mercado – e como ela se conecta ao papel do Engenheiro(a) de Dados no dia a dia. Você aprenderá o que é o Databricks, por que ele é tão utilizado pelas maiores empresas do mundo e como o profissional de Engenharia de Dados atua dentro desse ecossistema.

#### AULAS

##### **Aula 1 – O que é o Databricks?**

Entenda, de forma clara e prática, o que é o Databricks, sua arquitetura Lakehouse, como funciona o Delta Lake e por que a plataforma unifica Engenharia, Analytics, Data Science e IA em um único ambiente.

##### **Aula 2 – O que faz um Engenheiro(a) de Dados?**

Descubra o papel do Engenheiro de Dados dentro das empresas, suas principais responsabilidades, ferramentas utilizadas e como esse profissional se conecta às áreas de Data Science, BI e produtos digitais.



### SQL PARA ENGENHARIA DE DADOS

Este curso foca em SQL aplicado à Engenharia de Dados, indo além do básico para trabalhar diretamente com tratamento, limpeza e preparação de dados para análises, Data Warehouses e Data Lakes. Você aprenderá desde consultas fundamentais até técnicas de detecção de anomalias e tratamentos avançados de qualidade de dados.

#### AULAS

##### **Aula 1 – SQL, SELECT, WHERE, ORDER e JOIN**

Fundamentos das consultas SQL: seleção de dados, filtros, ordenação e junções entre tabelas.

##### **Aula 2 – Cálculos Básicos e Agregações**

Uso de funções de agregação (SUM, AVG, MIN, MAX, COUNT) e cálculos em colunas para gerar métricas.

##### **Aula 3 – Tratamento de Dados Duplicados**

Estratégias para identificar e remover duplicidades em tabelas com SQL.

##### **Aula 4 – Detecção de Anomalias**

Criação de regras e consultas para encontrar outliers e registros suspeitos na base.

##### **Aula 5 – Tratamento de Dados com CASE**

Uso do CASE WHEN para padronizar, corrigir e derivar novas informações a partir dos dados.

##### **Aula 6 – Casting**

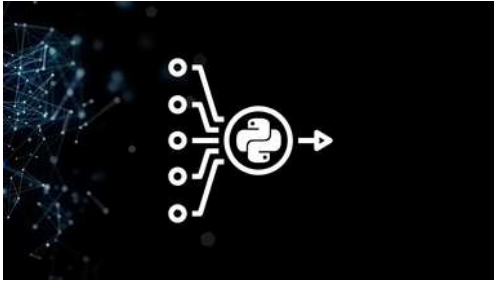
Conversão de tipos de dados (strings, números, datas) para garantir consistência e qualidade.

##### **Aula 7 – Tratamento de Dados Ausentes**

Identificação, substituição e tratamento de valores nulos ou ausentes.

##### **Aula 8 – Tratamento Final de Tabela**

Consolidação de todas as técnicas em um fluxo de preparação final da tabela para uso em relatórios, modelos e pipelines.



## PYTHON PARA ENGENHARIA DE DADOS

Este curso foi desenvolvido para ensinar Python aplicado diretamente ao dia a dia da Engenharia de Dados. Aqui, você aprenderá como carregar, transformar, limpar, padronizar e salvar dados em pipelines reais, integrando tudo com o Databricks. O foco é 100% prático, utilizando técnicas utilizadas por engenheiros em projetos profissionais.

### AULAS

#### **Aula 1 – Carregando Dados no Databricks com Python (CSV, MySQL e API)**

Aprenda diferentes formas de ingestão de dados e como conectá-los ao ambiente Databricks.

#### **Aula 2 – Analisando e Limpando Estruturas de Dados (cast, colunas e tipos)**

Entenda como validar estruturas, ajustar tipos e preparar dados para transformação.

#### **Aula 3 – Tratamento de Dados Duplicados com Python**

Técnicas práticas para identificar e remover duplicidades em datasets.

#### **Aula 4 – Tratamento de Dados Faltantes**

Estratégias para identificar, substituir ou descartar valores ausentes.

#### **Aula 5 – Substituição e Padronização de Valores**

Como corrigir, padronizar e normalizar valores inconsistentes ou categóricos.

#### **Aula 6 – Detecção e Tratamento de Anomalias**

Técnicas iniciais para encontrar outliers e comportamentos estranhos nos dados.

#### **Aula 7 – Manipulação de Datas e Criação de Colunas**

Como lidar com datas, timestamps e derivar novas colunas com Python.

#### **Aula 8 – Agregações, Merge e Salvando Dados no Databricks**

Transformações avançadas, junções de tabelas e escrita de dados tratados no Lakehouse.



## PYSPARK PARA ENGENHARIA DE DADOS

Este curso ensina, na prática, como utilizar o PySpark – a principal engine distribuída usada em projetos modernos de Engenharia de Dados. Você aprenderá a trabalhar com grandes volumes de dados, realizar transformações complexas, tratar inconsistências e construir um DataFrame final pronto para uso em Data Lakes, Data Warehouses e pipelines no Databricks.

### AULAS

#### **Aula 1 – Carregamento de Dados**

Como ler dados de diferentes fontes (CSV, Parquet, tabelas, storage) e criar DataFrames PySpark.

#### **Aula 2 – Select, Filter e Join**

Manipulação essencial de DataFrames: seleção de colunas, filtros condicionais e junções entre datasets.

#### **Aula 3 – Cálculos Básicos e Agregações**

Uso de funções agregadoras (SUM, AVG, COUNT, MIN, MAX) e cálculos derivados em colunas.

#### **Aula 4 – Tratamento de Dados Duplicados**

Estratégias para identificar e remover duplicidades em ambientes distribuídos.

#### **Aula 5 – Casting**

Conversão de tipos de dados para garantir consistência e integridade nos pipelines.

#### **Aula 6 – Detecção de Anomalias**

Identificação de outliers e comportamentos inesperados utilizando PySpark.

#### **Aula 7 – Tratamento de Dados Condicional**

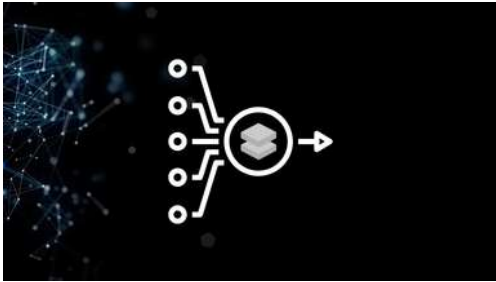
Uso de when, otherwise e expressões condicionais para padronização e correção de valores.

#### **Aula 8 – Preenchimento de Dados Ausentes**

Estratégias com fillna, imputações e validações de campos vazios.

#### **Aula 9 – Criando DataFrame Tratado**

Consolidação final: construção do DataFrame limpo e estruturado, pronto para gravação no Lakehouse ou uso em análises.



### DATABRICKS DATA LAKE HOUSE - ORQUESTRAÇÃO, CLUSTERS, CAMADAS, DATA QUALITY E POWER BI

Este curso mostra, na prática, como construir um Data Lakehouse completo dentro do Databricks – do carregamento inicial até a visualização final. Você aprenderá a configurar clusters, organizar o Lakehouse em camadas (Bronze, Prata e Ouro), criar orquestrações, aplicar regras de Data Quality e integrar o ambiente ao Power BI. É o curso perfeito para quem quer ver um fluxo end-to-end funcionando dentro do ecossistema Databricks.

#### AULAS

##### **Aula 1 – Introdução à Databricks**

Entenda a arquitetura do Databricks, o conceito de Lakehouse e como o Delta Lake transforma pipelines de dados.

##### **Aula 2 – Landing Zone: Criando e Conectando as Principais Bases no Data Lake**

Crie a camada inicial do Data Lake e conecte fontes externas, bancos, arquivos e sistemas ao ambiente.

##### **Aula 3 – Criação de Clusters e Orquestração**

Configure clusters otimizados, entenda os diferentes modos de execução e aprenda a criar orquestrações automatizadas com jobs e workflows.

##### **Aula 4 – Coding: Criação das Camadas Bronze, Prata e Ouro**

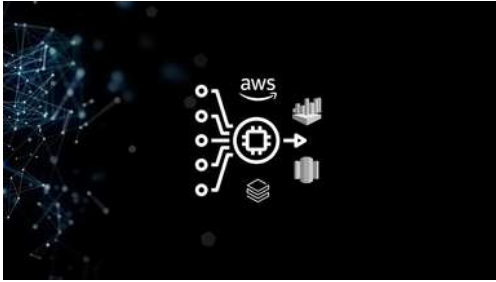
Construa o Lakehouse na prática, com transformações em PySpark, regras de limpeza e organização em múltiplas camadas.

##### **Aula 5 – Integração com o Power BI**

Aprenda a conectar o Power BI ao Databricks, criar dashboards e consumir dados diretamente do Lakehouse.

##### **Aula 6 – Data Quality**

Implementação de regras de qualidade, validações, checagens automáticas e criação de fluxos mais confiáveis para produção.



### AWS DATA LAKE HOUSE - AWS GLUE, AWS ATHENA E QUICKSIGHT

Aprenda a construir um Data Lake House completo dentro do ecossistema AWS, utilizando os principais serviços para ingestão, transformação, governança e visualização de dados.

Neste curso prático, você vai criar pipelines com o AWS Glue, consultar e estruturar dados no Athena, implementar Data Quality e entregar dashboards profissionais no QuickSight.

É um guia end-to-end para quem quer dominar Data Lake na AWS.

#### AULAS

##### **Aula 1 – Introdução ao Data Lake na AWS**

Entenda a arquitetura de Data Lake House, o papel do S3, Glue, Athena e como tudo se conecta dentro da AWS.

##### **Aula 2 – Conexões e Orquestração: AWS Glue e Gerenciamento de VPCs**

Criação de conexões seguras, permissões, VPCs, subnets e configuração da infraestrutura necessária para rodar pipelines.

##### **Aula 3 – Criando as Camadas do Data Lake com Glue + Athena**

Construção prática das camadas Bronze, Prata e Ouro utilizando Glue Jobs, Crawlers e consultas estruturadas com Athena.

##### **Aula 4 – Processo de Data Quality no AWS Glue**

Implementação de regras de qualidade, validações automáticas e checagens de integridade utilizando os recursos de Data Quality do Glue.

##### **Aula 5 – Dashboards e Integração com o AWS Data Lake: QuickSight**

Conecte o QuickSight ao Data Lake, crie visualizações e construa dashboards analíticos utilizando datasets criados no Athena.

##### **Aula 6 – Exclusão de Recursos na AWS**

Aprenda a encerrar ambientes de forma segura, removendo Glue Jobs, Crawlers, tabelas, buckets e outros recursos para evitar custos desnecessários.



### GCP DATA LAKE HOUSE - BIG QUERY, CLOUD STORAGE, DATAFLOW, CANVAS E LOOKER (DATA STUDIO)

Este curso ensina como construir um Data Lake House moderno utilizando todo o ecossistema do Google Cloud.

Você aprenderá desde a arquitetura base até pipelines reais com Dataflow, ingestão produtiva, camadas de tratamento no BigQuery, governança, qualidade de dados e visualização com Looker.

Também verá como usar o BigQuery Data Canvas, a nova interface com IA Generativa para análises e insights automáticos.

#### AULAS

##### **Aula 1 – Introdução ao GCP e Desenho da Arquitetura**

Conheça os principais serviços do Google Cloud, a estrutura de projetos e como desenhar a arquitetura de um Data Lake House robusto.

##### **Aula 2 – Redes, Conexões e Criação de VM no GCP**

Entenda VPC, subnets, firewall, criação de máquinas virtuais e como conectar o ambiente a bases de dados externas.

##### **Aula 3 – Orquestração com Dataflow, Cloud Storage e BigQuery Data Transfer (Bases Produtivas)**

Utilize pipelines reais para ingerir dados produtivos via Dataflow e integrá-los ao Cloud Storage e BigQuery.

##### **Aula 4 – Orquestração com Dataflow e BigQuery Data Transfer usando Cloud Storage (Mock)**

Crie pipelines simulados (mock) para testar arquiteturas, automatizar ingestão e validar fluxos de dados.

##### **Aula 5 – Modelagem e Criação das Camadas RAW → TREATED → REFINED no BigQuery**

Estruture seu Lake House no BigQuery com camadas de ingestão, transformação e refinamento, seguindo boas práticas de engenharia.

##### **Aula 6 – Governança, Metadados e Data Quality no BigQuery (Data Profile + Data Quality Rules)**

Aprenda como implementar regras de qualidade, analisar metadados, perfis de dados e aplicar validações diretamente no BigQuery.

##### **Aula 7 – Data Canvas no BigQuery, IA Generativa e Integração com Looker**

Explore o BigQuery Data Canvas, gere análises com IA Generativa e conecte datasets ao Looker (Data Studio) para criar dashboards.



### AZURE DATA LAKE HOUSE - DATABRICKS, AZURE DATA FACTORY & AZURE PURVIEW

Este é um dos cursos mais completos da formação, levando você pela construção de um Data Lake House Enterprise na Azure do zero até um ambiente totalmente governado, monitorado e preparado para produção.

Ao longo das aulas, você aprenderá a criar toda a arquitetura de dados utilizando Databricks, Azure Data Factory, Azure Purview, Key Vault, Cosmos DB, SQL Server, MySQL, Blob Storage, além de trabalhar com múltiplas zonas do Lakehouse – Landing, Raw, Dev, Work/Trusted, Sensitive e Refined.

Também vai implementar orquestrações, pipelines D-1, controle de custos, modelagem, governança, segurança, linhagem, Data Quality e consumo analítico.

É um guia end-to-end que simula a realidade das grandes empresas – ideal para quem quer dominar engenharia de dados moderna na Azure.

#### AULAS

##### **Módulo 1 – Fundamentos e Arquitetura**

###### **Aulas 1 a 3**

Compreensão dos dados, comparação entre clouds e desenho da arquitetura oficial do Lakehouse.

##### **Módulo 2 – Criação de Bases e Serviços de Dados**

###### **Aulas 4 a 6**

Criação de bancos SQL, MySQL e NoSQL dentro da Azure.

##### **Módulo 3 – Databricks na Prática**

###### **Aulas 7 a 14**

Provisionamento do workspace, cofre de senhas, clusters, conexão segura e criação dos primeiros pipelines D-1.

##### **Módulo 4 – Azure Data Factory & Orquestrações**

###### **Aulas 15 a 25**

Integrações com múltiplas bases, pipelines resilientes, retries e ingestão para o Data Lake.

##### **Módulo 5 – Raw Zone**

###### **Aulas 26 a 32**

Montagem da camada raw, organização de containers, bancos e notebooks.

##### **Módulo 6 – Criação de Tabelas Raw**

###### **Aulas 33 a 45**

Ingestão completa de SQL Server, MySQL, NoSQL, APIs e dados externos.

## **Módulo 7 – Dev Zone**

### **Aulas 46 a 57**

Configuração da área de desenvolvimento, quotas, clusters e tratamento inicial de dados com PySpark.

## **Módulo 8 – Trusted / Work Zone**

### **Aulas 58 a 66**

Refinamento, tabelas tratadas, controles de governança e preparação para produção.

## **Módulo 9 – Governança com Azure Purview**

### **Aulas 67 a 76**

Mapeamento de dados, linhagem, coleções, segurança e permissionamento.

## **Módulo 10 – Sensitive Zone**

### **Aulas 77 a 85**

Identificação de dados sensíveis, proteção de acessos, tabelas restritas e segurança avançada.

## **Módulo 11 – Refined Zone**

### **Aulas 86 a 90**

Modelagem refinada, permissões, estrutura final de tabelas e cluster especializado.

## **Módulo 12 – Data Warehouse Zone**

### **Aulas 91 a 95**

Exportação para bases relacionais, criação de esquemas de DW e integração com Power BI.

## **Módulo 13 – Monitoramento & Encerramento**

### **Aulas 96 a 97**

Monitoramento do Databricks, ADF e Purview, além de considerações finais e próximos passos.

## CONTATO

---

✉ contato@rodolfomoreira.com.br

📞 11 9 1600-4334

🌐 <https://www.linkedin.com/in/rodolfo Rodriguesmoreira/>