

Open Class

Engenharia de Dados

com André Sionek
Rhuan Lima
Rafael Dias



_Quem é o André?

- Engenheiro de dados na [Skyscanner](#)
- Curitibano, mas mora em Londres
- **Não se formou** em Física, administração, engenharia, ou matemática
- Trabalhou na Polyteck, HSBC, Grupo Boticário, Olist, EmCasa, Gousto, Skyscanner e agora na Util
- Escreve no Medium e é instrutor em bootcamps com certa regularidade



[Como vim parar em Londres?](#)

Casa Raulino Kretzer - arthur.raulino.kretzer@gmail.com - IP: 189.4.77.203

[Como comecei a carreira de engenheiro de dados?](#)

_Quem é o Rhuan?

- Especialista em Engenharia de Dados
- Curitibano, mas mora em Curitiba
- **Não se formou** em BSI, **mas se formou** em Estatística pela UFPR
- Atualmente trabalha no Grupo Boticário mas teve passagem em empresas como HSBC e Bradesco
- Pregador do Python e agora fazendo parte do grupo de instrutores de bootcamps



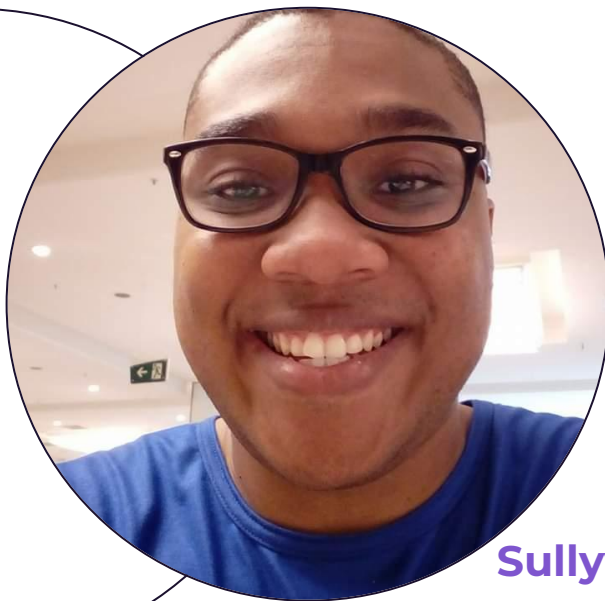
_Quem é o Rafael?

- Cloud Engineer - AI & ML na [Oracle Cloud](#)
- Curitibano, morando em Curitiba
- **Formado** em Estatística pela UFPR, e Especialista em Data Science & Big Data pela UFPR e também pela Correlation One
- Trabalhou no Grupo Boticário, MadeiraMadeira, Pipefy e agora está na Oracle Cloud
- Professor de Data Science e Big Data, e agora também de Engenharia de Dados



```
from how import  
    como_funcionara_o_bootcamp
```

Apresentação



Sullyvan Madeira

Analista de CX

sullyvan@howedu.com.br



***Somos a How Education,
Edtech brasileira pioneira
em bootcamps cohorts.***

***Impulsionamos a sua
carreira em áreas de alta
demanda.***



how

Mentor da Comunidade



Rodrigo Azevedo

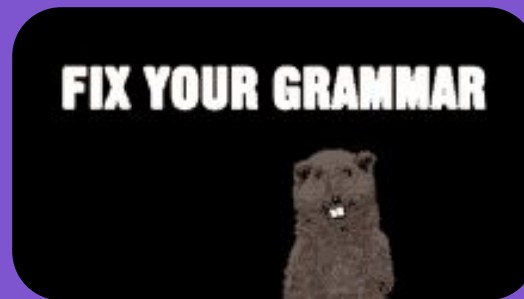
Data Engineer

Lopes Consultoria de Imóveis

how

Vocabulário How

- ☐ Bootcamp eurse
- ☐ Facilitador(a) professor(a)
- ☐ Encontro aula
- ☐ Módulos e episódios
- ☐ Cohort



Corrija seu vocabulário



O que é Cohort?

Significa grupo!

Nossa comunidade é no Discord!

Aqui

Você vai encontrar as
informações sobre o
que vai acontecer ao
longo dessas 10
semanas de bootcamp!

Bootcamp Engenharia de Dados

Com **Rafael Roberto Dias**



how

Serão 10 semanas intensas

A cada uma delas, você terá que

- Assistir os episódios gravados
- Participar dos encontros ao vivo (ou assistir gravado)
- Realizar as atividades propostas
- Organizar dúvidas
- Manter a energia lá em cima!

Então organize seu tempo porque você vai precisar 😊



Encontros ao vivo

- ❖ **Serão 11 encontros que acontecerão toda terça-feira, das 19h30 às 22h.**



Módulos

- **Conteúdos Extras** - Materiais Complementares de outras edições.
- **Introdução à Engenharia.**
- **Fundamentos de Ingestão de Dados.**
- **SQL.**
- **Capturando Dados de uma API.**
- **Capturando Dados com Crawlers.**
- **Testes + Jenkins.**
- **Git + GitHub.**

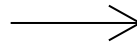
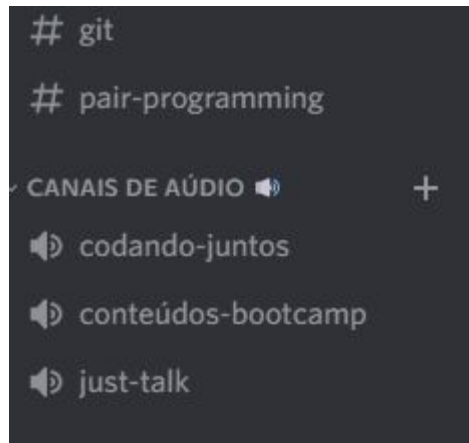
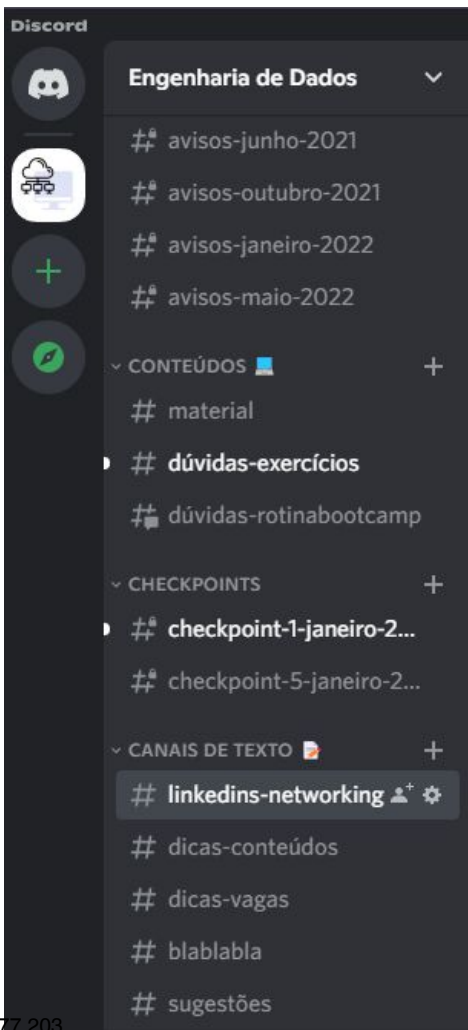
Módulos

- Terminal Vscode.
- Introdução à AWS.
- Integração Contínua.
- Data Lakes.
- AWS Glue + Athena.
- Redshift + Spectrum.
- Apache Airflow.
- Terraform e Cloudformation.



Ferramentas

Discord | Canais e Avisos



how

Ferramentas

Memberkit | Avaliações e % de conclusão



Ferramentas

Memberkit | Avaliações e % de conclusão

▼ Introdução à Engenharia de Dados ...		
▶ How Bootcamps - Engenharia de Dados Teaser - 01:00	4.9	★
▶ Introdução ao bootcamp - 09:13	4.9	★
▶ Hora da Prática: Loterias parte 01 - 14:15	4.8	★
▶ Hora da Prática: Loterias atualização - 16:14		☆
▶ Hora da Prática: Loterias parte 02 - 18:52	4.7	★
▶ Hora da prática: Loterias parte 03 - 15:01	4.7	★
▶ Módulos e Conteúdos - 14:22	4.8	★

☰

💡

<

>

0%

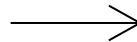
Avaliação

☆☆☆☆☆

Marcar como concluído

Ferramentas

Zoom | Link é sempre o mesmo



how

Certificado

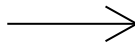
70% do
conteúdo
gravado
assistido

3

CheckPoints

**momentos de interação
que serão estabelecidos
como requisitos para o
certificado**

estarão explícitos na
descrição do Memberkit
e nos avisos



**1 ano para
finalizar!**

how

Checklist inicial

- ☐ Encontros adicionados à agenda
- ☐ Acesso ao Discord
- ☐ Check no canal de avisos
- ☐ Acesso ao Memberkit
- ☐ Atualizar o perfil no Memberkit
- ☐ Assistir aos vídeos introdutórios



Thanks!



how

how

_Qual problema um
engenheiro de dados resolve?



_Por que o problema existe?



_O que o engenheiro de dados faz?



```
from how import  
    engenharia_vs_ciencia_vs_bi
```

_Engenheiro de Dados

_Cientista de Dados

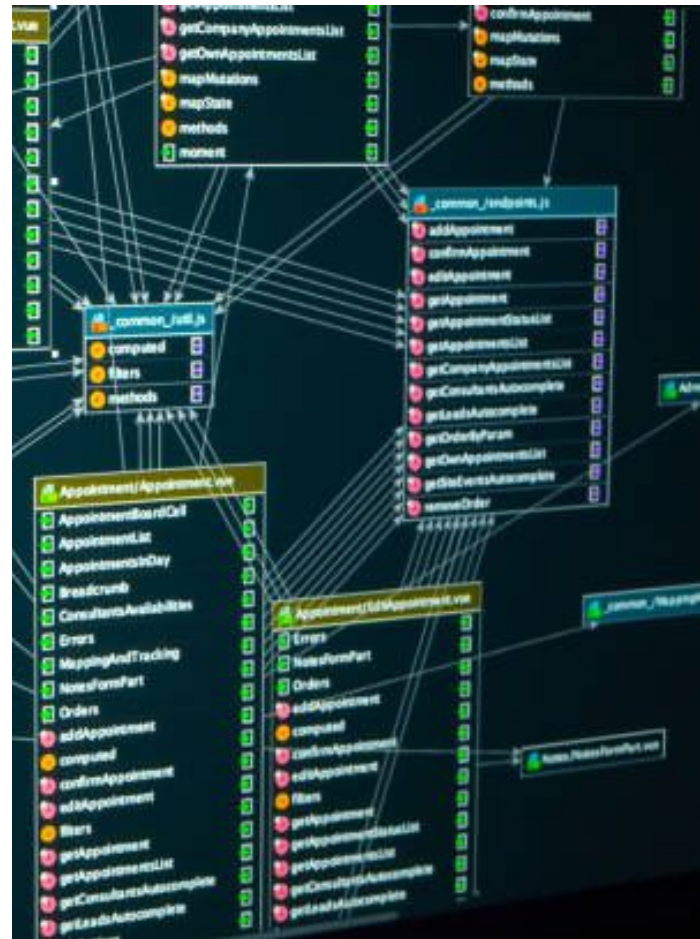
_Analista de Business Intelligence

```
from how import  
    areas_do_engenheiro_de_dados
```

_Infraestrutura/Plataforma

Responsável por criar e manter a infraestrutura de dados da empresa

- DataOps - Evolução do DBA
- Responsável pela estrutura de dados: Data Lake, Warehouse, Lakehouse
- Serviços de Streaming
- Pipelines de testes automatizados e de produção
- Facilitar o trabalho do time de dados
- Garantir boas práticas na utilização da arquitetura de dados
- Exemplos de ferramentas: Terraform, Cloudformation, Kafka, Cloudwatch



_Criação e Gestão de Pipelines

Responsável por criar e manter o fluxo de dados das fontes para as áreas clientes

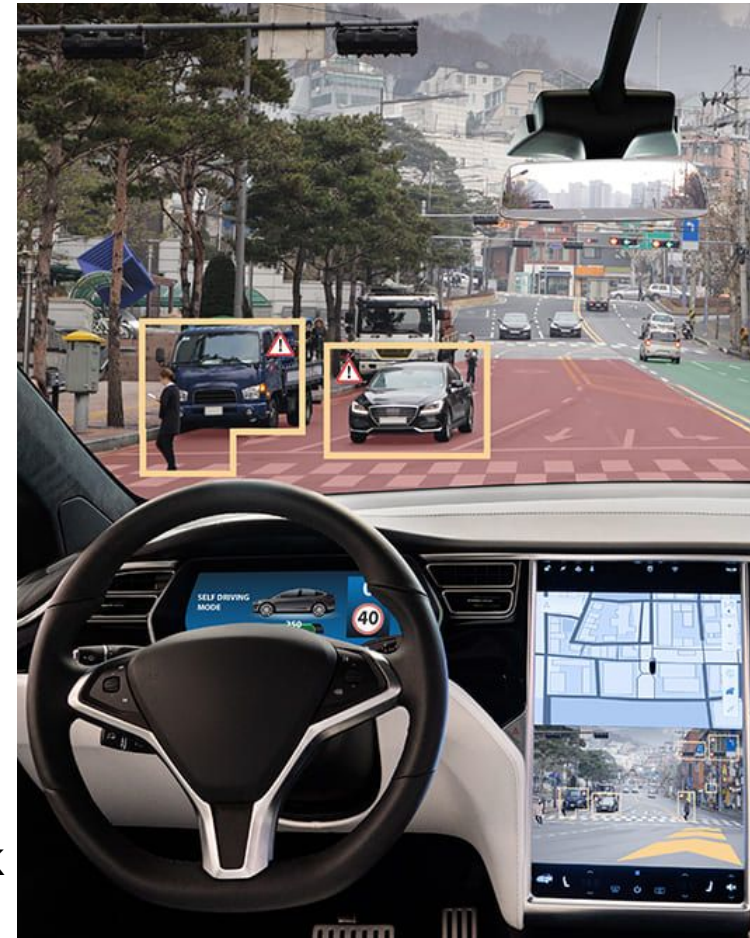
- Área mais demandada no momento
- Criação de scripts para extração de dados por API e DB
- Criação e manutenção de crawlers para extração de dados diretamente da web
- Criação e gerenciamento de DAGs no Airflow
- Realizar processos de ETL e higienização dos dados
- Exemplos de ferramentas: Airflow, Spark, Pandas, SQL, Lambda, Buckets



_MLOps

Responsável por colocar modelos de AI & ML em produção e acompanhar a qualidade destes

- Um dos melhores amigos de um Cientista de Dados
- Realiza o deploy e acompanhamento dos modelos
- Cria pipeline para envio e recebimento de requisições
- Monitora o desempenho dos modelos
- Exemplo de ferramentas: Docker, Kubernetes, APIs' & Endpoints', SQL, Python, R, Java, Spark



Analytics

Responsável por aplicar regras de negócios e criar modelos de dados para serem consultados pelo negócio

- Trabalha muito próximo do analista de dados ou de BI
- Entende as peculiaridades do negócio: KPI, OKR, MC1, MC2
- Unifica dados de diferentes fontes para criar visão única
- Cria modelos de dados com fatos e dimensões
- Garante correta atualização dos dados e cálculos precisos das regras de negócio
- Exemplo de ferramentas: SQL, NoSQL, DBT, Airflow, Tableau, Power BI, Python, R



_Pasteleiro

Responsável por todas as funções de engenharia de dados

- Seria engraçado, senão fosse a mais pura realidade



```
from how import  
    junior_pleno_senior
```

_Diferenças



_Diferenças



Engenheiro de Dados Júnior

- Já aguenta bastante porrada
- Consegue manter os pipelines existentes
- Constrói ingestões de dados mais simples
- Trabalha diretamente com pessoal da área
- Está sendo desenvolvido pelo Pleno

_Diferenças



Engenheiro de Dados Pleno

- Consegue botar o Vegeta pra correr
- Implementa novos pipelines de dados
- Constrói ingestões de dados batch
- Trabalha com pessoal da área e tem contato com a área de negócios
- Propõe novas maneiras de ingestão de dados ao Sênior

_Diferenças



Engenheiro de Dados Sênior

- Não derrota o **Superman** por não ser da DC
- Cria e implementa novos pipelines de dados
- Constrói ingestões de dados batch e stream
- Trabalha com pessoal da área e com a área de negócios
- Propõe e implementa novas arquiteturas de gestão de dados em Multicloud

```
from how import  
    outros_recursos_para_aprender
```


Programação / Python

- Codecademy
- Alura
- Coursera
 - [Fundamentals of Computing Specialization](#)
- Udemy
 - [Design Patterns in Python](#)

Cloud Infrastructure:

Oracle Cloud

[Oracle OCI Training and Certification](#)

GCP

[Google Cloud Courses and Training](#)

AWS

[AWS Training - Free Cloud Training](#)

Azure

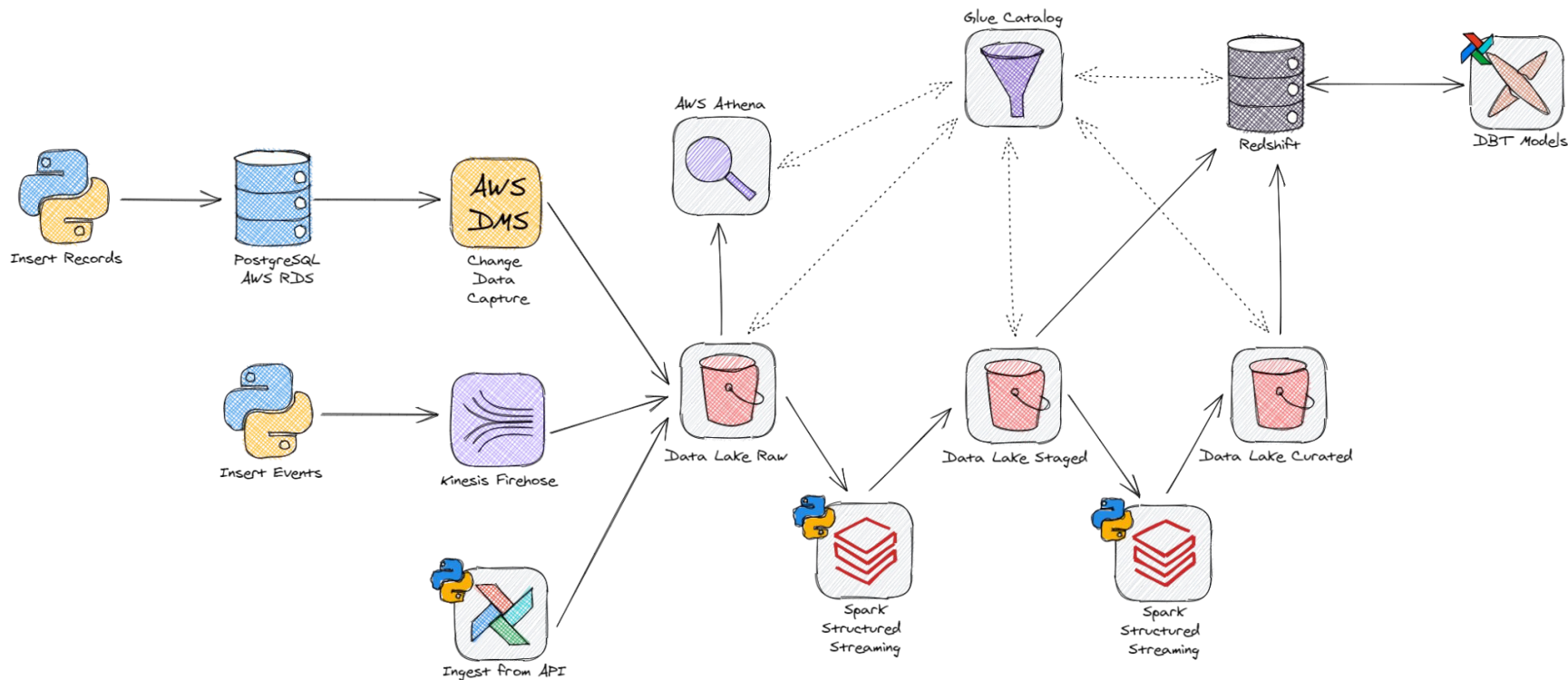
[Microsoft Certified: Azure Fundamentals - Learn](#)

_Spark

- Udemy
- Alura
- Coursera
- Databricks

```
from how import  
esperado_ao_termino_deste_bootcamp
```

_Vcs serão capazes de construir ao fim do bootcamp



_Demo Ingestão de Dados

- Ingerir eventos com Kinesis Firehose
- Salvar no S3
- Catalogar os dados com Glue
- Realizar consultas com Athena

