

Introdução à Engenharia de Dados

com André Sionek e Rhuan Lima

```
$cd modulo_2
```

```
$. /fundamentos_de_engenharia_de_dados
```

```
$_
```

```
from how import  
boas_vindas
```

_Quem é o André?

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

[Como vim parar em Londres?](#)

[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 a How?

how

Fale conosco

HOW Bootcamps

Construa novas habilidades com bootcamps ao vivo, práticos e de curta duração.

Você aprenderá com facilitadores do mundo real que trabalham em empresas como Nubank, Hotmart, EBANX e 99.

Bootcamps nas áreas de UX Design, Product Management, Customer Success e Dados

#howbootcamps #skillsnotdegrees



Bootcamps

Q Pesquisar

III Filtros

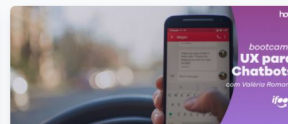


Workshop UX Design para não Designers com Vivi Tavares (Gonddo)

Online
Edição Junho | Noturno

DATA
JUN
15
23:00

SAIBA MAIS



UX para Chatbots com Valéria Romano (iFood)

Online
Edição Junho | Noturno

DATA
JUN
22
23:30

SAIBA MAIS



Bootcamp Customer Success

Online
Julho | 10 semanas | Noturno

DATA
JUL
07
23:30

SAIBA MAIS



Acessibilidade Digital



UX Writing



UX para Chatbots

_Informações Importantes

- Agenda
- Pair/Group Programming
- Lives + Aulas Gravadas
- Memberkit
- Canal no Discord
- LinkedIn
- GitHub/GitLab

_junho

D	S	T	Q	Q	S	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

10 | Encontro Inaugural

17 | Live com convidado

24 | Prática e cases

29 | Live com convidado

_julho

D	S	T	Q	Q	S	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

01 | Prática e Cases

06 | Group programming

08 | Prática e Cases

13 | Live com convidado

15 | Prática e Cases

20 | Group Programming

22 | Prática e Cases

27 | Group Programming

29 | Prática e Cases

_agosto

D	S	T	Q	Q	S	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

10 | Encontro Inaugural

17 | Live com convidado

24 | Prática e cases

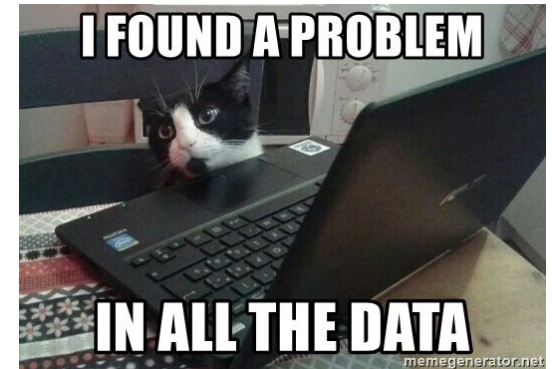
29 | Live com convidado

```
from how import  
    o_que_e_engenharia_de_dados
```

_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_de_dados_vs_bi
```

_Cientista de Dados

_Engenheiro de Dados

_Analista de Business Intelligence


```
from how import  
    tipos_de_engenheiros_de_dados
```

_Infraestrutura/Plataforma

Responsável por criar e manter a infraestrutura de dados da empresa.

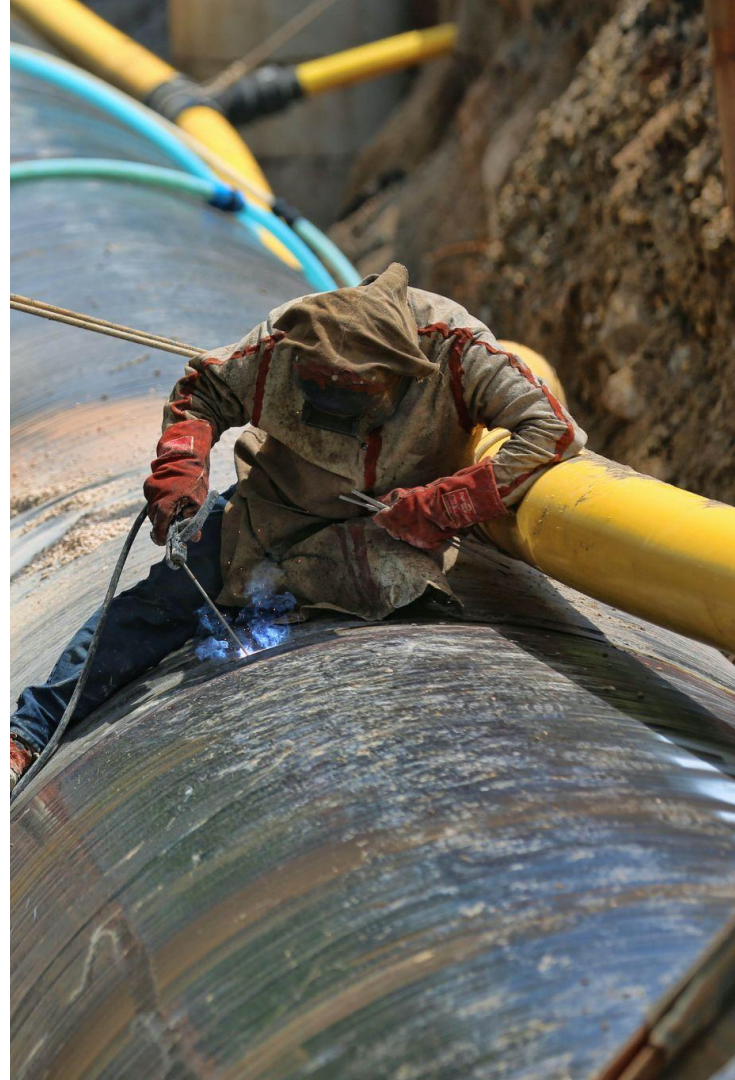
- DataOps / Evolução do DBA
- Cuida do data lake e data warehouse
- Serviços de streaming
- Pipelines de deploy e testes automatizados
- Cria ferramentas para facilitar o dia a dia do restante do time de dados
- Guardião de boas práticas e da arquitetura
- Ferramentas: Cloudformation, Kafka, PagerDuty, Cloudwatch...



_Pipelines

Responsável por criar e manter o fluxo de dados das fontes para os consumidores.

- Tipo mais comum de profissional
- Cria scripts para extrair dados de APIs e DBs
- Cria crawlers para extrair dados de páginas da web
- Cria e gerencia DAGs no Airflow
- Transforma e limpa os dados
- Ferramentas: Airflow, Spark, Pandas, SQL, Lambda



_Machine Learning

Responsável por colocar e manter modelos de machine learning em produção.

- Braço direito do cientista de dados
- Cria pipelines para transformar e criar variáveis
- Faz o deploy de modelos de ML
- Monitora o desempenho dos modelos
- Ferramentas: Pandas, TensorFlow, SageMaker, SQL, MLFlow, Spark, Docker, Lambda



_Analytics

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

- Trabalha próximo do analista de dados/BI
- Junta dados de diferentes fontes para criar uma visão unificada
- Cria modelos de dados com fatos e dimensões
- Entende as peculiaridades do negócio
- Ferramentas: SQL, DBT, Airflow, Ferramentas de BI



_Severino



```
from how import  
    outros_recursos_para_aprender
```

Programação / Python

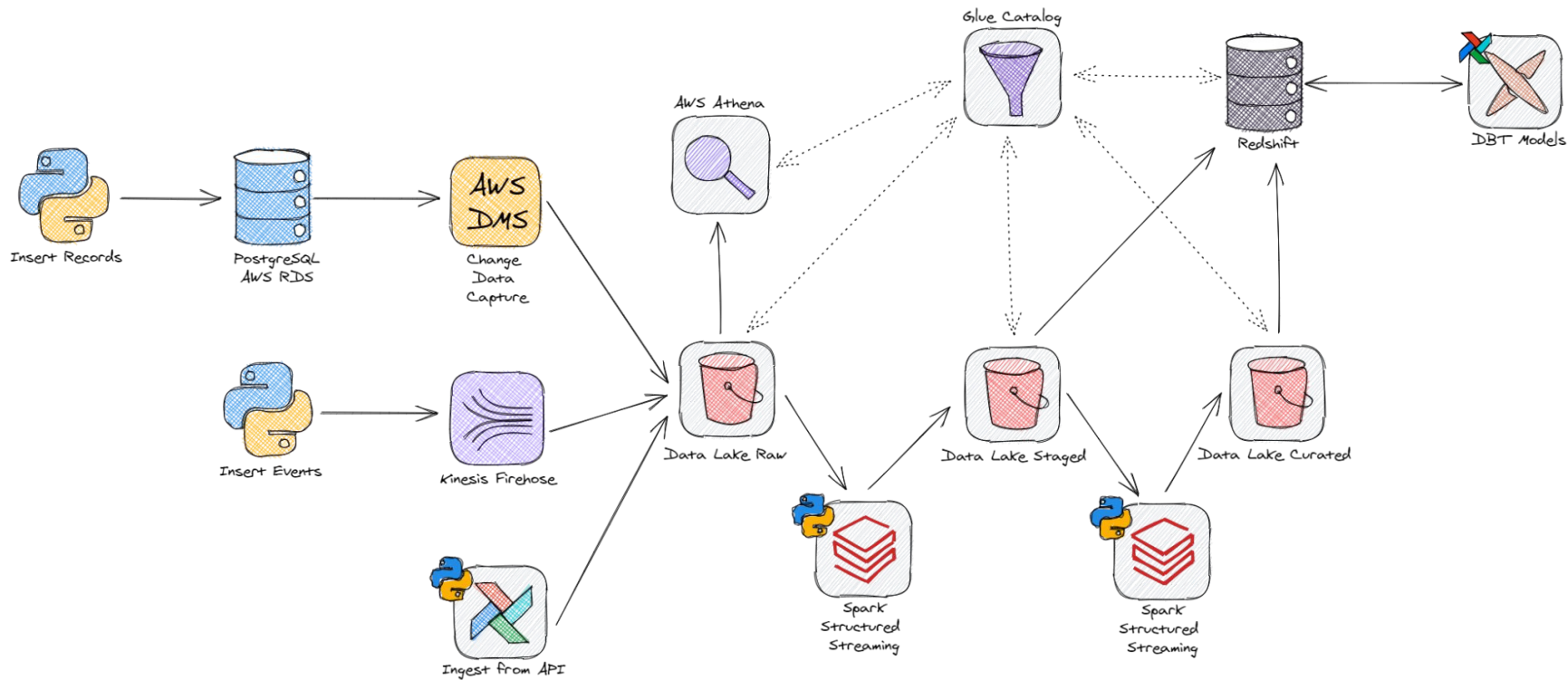
- Codecademy
- Alura
- Coursera
 - [Fundamentals of Computing Specialization](#)
- Udemy
 - [Design Patterns in Python](#)

AWS

- Alura
- Udemy
- [AWS Training](#)

Spark

- Alura
- Udemy
- Databricks



Time to code...

_Exercício

Ingerir eventos com Kinesis Firehose, salvar no S3 e fazer query com Athena.

