

## Bootcamp: Engenheiro(a) de Dados (Cloud)

### Trabalho Prático

#### Módulo 4: Processamento de fluxo contínuo de dados

#### Objetivos de Ensino

Exercitar os seguintes conceitos trabalhados no Módulo:

1. Conceitos específicos das pipelines em real time.
2. Arquitetura do Kafka.
3. Migração de Dados utilizando Kafka Connect.

#### Enunciado

Você é Engenheiro de Dados da Monsters Inc., uma loja virtual de artigos *nerd* que está super em alta e com um fluxo enorme de novos clientes. Os executivos da Monsters Inc. já entenderam que contabilizar cadastros de novos clientes uma vez por dia não é suficiente para atendê-los com a velocidade que eles precisam e, portanto, decidiram implementar um pipeline em real time.

Você será responsável por implementar um pipeline em tempo real que faz a aquisição de novos cadastros de clientes numa database relacional e os escreve em um *data lake* no Amazon S3.

## Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Clonar o repositório git  
[https://github.com/neylsoncrepalde/igti\\_edc\\_mod2\\_aula\\_interativa\\_1](https://github.com/neylsoncrepalde/igti_edc_mod2_aula_interativa_1)
2. Executar as instruções do exercício dentro da pasta connect até o passo 7 - Registrar os parâmetros de configuração do connector no Kafka.
3. Responder às perguntas do questionário com base na atividade.