

Bootcamp IGTI: Engenheiro(a) de Dados Cloud

Trabalho Prático

Módulo 2	Processamento e Fluxo Contínuos de Dados
-----------------	---

Objetivos

- ✓ Utilizar os principais serviços de nuvem para Engenharia de Dados;
- ✓ Utilizar o Kubernetes como Gerenciador de Containers;
- ✓ Ingerir dados em *real time* no Apache Kafka;
- ✓ Implementar um pipeline de processamento de Big Data em *real time*;
- ✓ Realizar Processamento de dados utilizando Ksqldb;
- ✓ Disponibilizar dados no Apache Pinot.

Enunciado

Você é Engenheiro(a) de Dados e precisa realizar a migração de uma tabela relacionada a clientes, que é altamente requisitada pela área de negócios para, assim, realizar análises comportamentais dos clientes. O gestor de sua área iniciou um projeto de migração para que esses dados sejam disponibilizados em tempo real, e você será responsável por construir uma pipeline em *real time* para disponibilizar esses dados no DW. Você precisará realizar o processamento utilizando ferramenta adequada e disponibilizar o dado para consultas dos usuários de negócios e analistas de BI.

Para a realização dessa atividade, recomenda-se o uso dos serviços AWS.

Atividades

Você deverá desempenhar às seguintes atividades:

1. Crie um RDS Postgres que tenha acesso público liberado;
2. Clonar o projeto <https://github.com/carlosbpy/igti-k8s-exercise>;
3. Ingerir dados da tabela de customers para o postgres;
4. Deployar EKS na AWS ou utilizar o Minikube;
5. Migrar tabela de customers para a estrutura do Kafka;
6. Realizar o processamento dos dados da tabela de customers utilizando o KsqlDB;
7. Disponibilizar dados da tabela de customers em *real time* para o Apache Pinot.