



escola
britânica de
artes criativas
& tecnologia

Profissão: Analista de dados



4º PROJETO: PIPELINE DE DADOS DO TELEGRAM



GUIA DA AULA 7



Crie o Storytelling



Acompanhe aqui
os temas que
serão tratados
na videoaula

Imagens

Notebook

GitHub

Melhorias

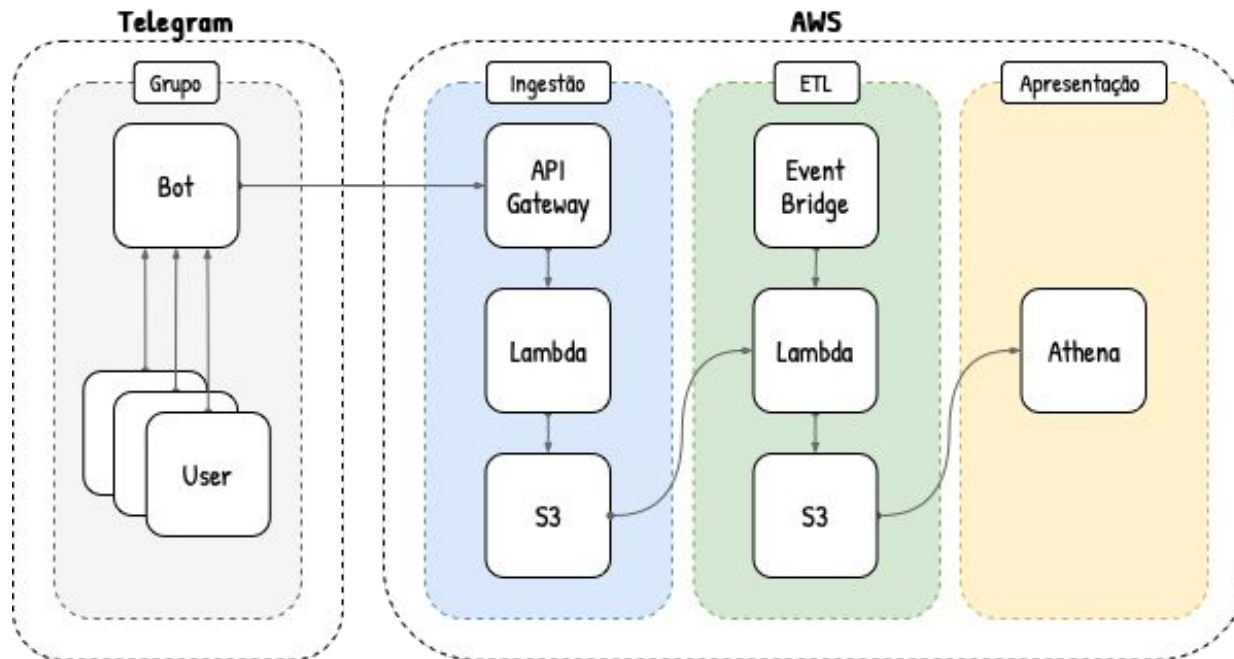


1. Imagens

Imagens são adequadas para o resumo e a explicação de assuntos complexos, como a arquitetura que desenvolvemos. Afinal, o humano é muito bom no processamento de dados não estruturados. Utilize a imagem fornecida para divulgar o seu projeto em redes sociais, como o LinkedIn. Não se esqueça de me enviar ou me marcar (*link* para o meu perfil <https://www.linkedin.com/in/andremarcosperez/>), ficarei muito feliz em comentar e compartilhar a publicação.



- Arquitetura



2. Notebook

A criação de um notebook e sua implantação na plataforma do Kaggle (<https://www.kaggle.com>) é umas das maneiras mais dinâmicas de expor o seu trabalho. Como sugestão, o notebook não precisa conter todos os detalhes como os *notebooks* do módulo 43 e 44 e pode apresentar a seguinte estrutura:

- **Contexto**
 - Posicionar o leitor no contexto de *chatbots*;
 - Posicionar o leitor sobre a diferença entre dados transacionais e analíticos que, por sua vez, justificam a criação do projeto.



- **Arquitetura**

- **Sistema transacional:** comentar brevemente sobre o Telegram e sua API de *bots*;
- **Sistema analítico:** apresentar com códigos as três etapas: ingestão, ETL e apresentação.



- **Análise Exploratória de Dados**

- **Fonte:** Analisar o dado com ele é produzido pela API de *bots* do **Telegram** através do uso do método *getUpdates*;
- **Destino:** Analisar o dado na etapa de apresentação com consultas SQL e potenciais visualizações sobre os dados retornados.

Dica: Foque bastante na análise exploratória de dados e mencione o que foi realizado nas demais etapas.



3. GitHub

Uma excelente opção é a criação de um repositório do **GitHub** para armazenar todos os códigos gerados no projeto:

- códigos Python das funções do `AWS Lambda` das etapas de ingestão e ETL;
- códigos SQL da etapa de apresentação do `AWS Athena`.

Outra dica é fazer o *commit* das imagens no repositório e utilizá-las no arquivo `README.md`. Você pode utilizar o repositório no link <https://github.com/andre-marcos-perez/data-pipeline-demo> como exemplo, ele contém uma versão deste projeto que apresentei ao vivo (link da apresentação: <https://www.youtube.com/watch?v=sVFT-XK6rk4>).



4. Melhorias

- **Infraestrutura**

A tabela do `AWS Athena` contém atualização automática de novas partições. Uma melhoria seria adicionar essa funcionalidade no `AWS Lambda`.



- **Dados**

No projeto, estamos transformando e apresentando apenas os dados referentes a mensagens de texto. Contudo, estamos ingerindo todas as mensagens que são enviadas ao grupo do **Telegram**: vídeo, áudio, arquivos, etc. Uma melhoria seria adicionar novos campos na função do `AWS Lambda` da etapa de ETL e na tabela do `AWS Athena` da etapa de ingestão, ou ainda criar uma novas funções de ETL e tabelas de apresentação.



- **Consultas**

A tabela do `AWS Athena` da camada de apresentação contém dados que apresentam diversas oportunidades de criação de novas consultas analíticas. Uma melhoria seria a criação de novas consultas SQL..

