Estruturas de pastas do projeto:

Path:

mnt\dm2\document-streaming-main

Airflow:

* Script para criar a Dag.

ApacheSpark:

* Onde fica salvo os códigos de pyspark por meio de Jupyter Notebook, todo o processo de ETL é feito dentro dos arquivos desta pasta.

API-Ingest:

* Onde fica salvo os códigos de criar o kafka consulmer, as configurações de Postman, e a configuração de API Ingest para criar a base de envio principal do Ingest.

Archive:

* Arquivos outdated ou que não está mais em uso.

Client:

* Onde cria as bases Dummy e scrpts.

Dags:

* Onde fica configurado todas as Dags do Airflow e estruturas de como elas são montadas.

Database:

* Criação das informações dummy e envio das tabelas e dataframes para o Postgresql.

Documentacoes:

* Onde ficam salvas as documentações e arquivos para analise da IA de documentação.

Dummy\_tables\_creates:

* Dummy create tables de tabelas para alimentar a IA.

Fbgemm:

* Arquivo para analise do pytorch e speech to text.

Libs:

* Onde estão as libs do projeto.

Logs:

* Cadastro de logs e analises.

Loghash:

* Onde fica salvo as configurações de loghash para analises e links dos containers.

Scripts:

* Scripts para acesso ao airflow com os containers docker.

Streamlit:

* Onde ficam salvos todos os streamlits e configurações das páginas de streamlit.