

**2)A**

A estrutura SQL é apropriada, pois os dados dos arquivos são organizados de forma estruturada e claramente definidos por seus relacionamentos. As tabelas seguem o conceito de schema, garantindo uma organização consistente e bem definida

**2)B**

Os dados passaram pelo processo de ingestão e importados para o Lake, um código Python é responsável pela extração dos dados dos arquivos

**"ingest.py"**

Foi feita uma pequena simulação de uma arquitetura medalhão, então o Lake possui duas camadas, Bronze e Gold (em um ambiente real também existiria a camada Silver e em alguns casos uma Raw, mas para esse simulado apenas essas duas foram criadas)

BRONZE - Os dados são armazenados de forma bruta, sem qualquer tipo de transformação ou tratamento.

* Script SQL: "**BRONZE.TBL\_ATRACACAO.SQL" "BRONZE.TBL\_CARGA.SQL”"**

GOLD - Nessa camada os dados são armazenados após já receberem as transformações necessárias, seja etapas de junções, transformações de colunas, renomeações etc.

* Script SQL: "**GOLD.TBL\_ATRACACAO.SQL" "GOLD.TBL\_CARGA.SQL”"**

Para o tratamento dos dados entre as camadas, PySpark foi utilizado. Transformação de coluna do tipo String para DateTime foi uma das etapas aplicadas.

**"transform.py"**

**2)C)**

Script da consulta **“QueryExcel.sql”.**

**3) Airflow**

Foi criado a DAG no Apache Airflow.   
Script DAG: **“antaq\_fiec.py”**