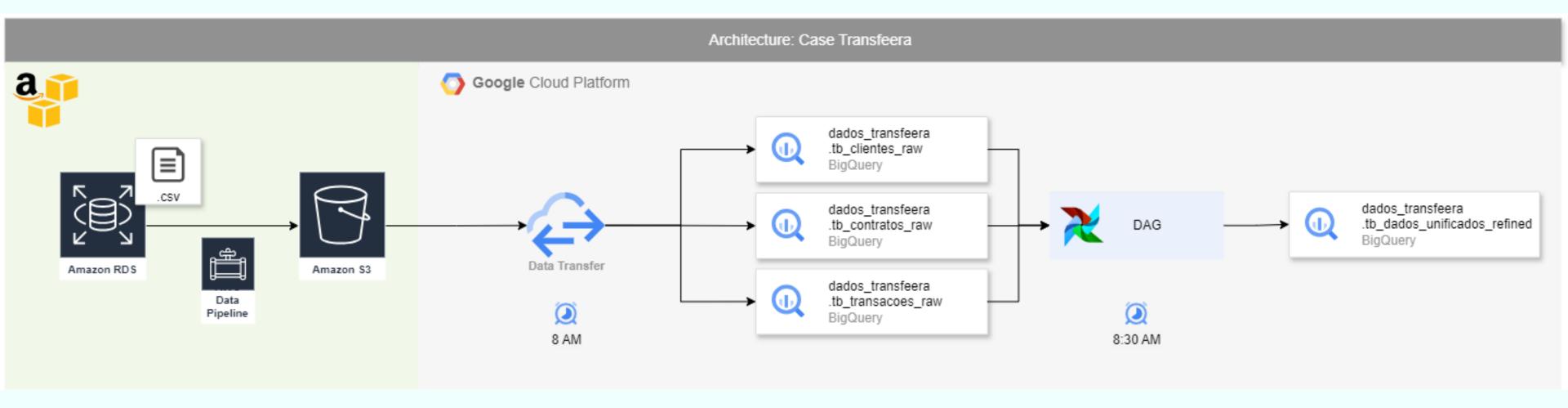


Case Analytics Engineer

Amanda Louise Costa Nascimento

Arquitetura Proposta





Motivação

- Facilidade de manutenção
- Dados atualizados diariamente, fora de horários de pico e pouco antes do início do expediênte



- Desenvolvimento de solução na AWS e GCP para carga dos csv's
- Utilização de ambiente trial da GCP para crição e armazenamento de tabelas
 - Foram criadas 3 tabelas raw utilizando os arquivos disponibilizados
- DAG dag_case_transfeera.py criada
 - Exceptions em cada função
 - Deduplicação de dados
 - Upper case em campos string
 - Cruzamento de tabelas
- Criação de tabela refined para conexão com dashboard
 - Particionada por data da movimentação



- Criação de campo timestamp de data de ingestão para evitar processamento duplicado. Apenas o delta é carregado.
- Desenvolvimento de dashboard com possíveis análises

OBS: Para esse case, foi considerado que os arquivos possuem nomes e estruturas fixas. No caso de arquivos com nomes diferentes mas de mesmo prefixo, seria possível parametrizar os nomes.



• Simulação de conexão entre BigQuery e S3 utilizando o Data Transfer. Nele, é possível realizar o agendamento das cargas, permitindo atualização diária.

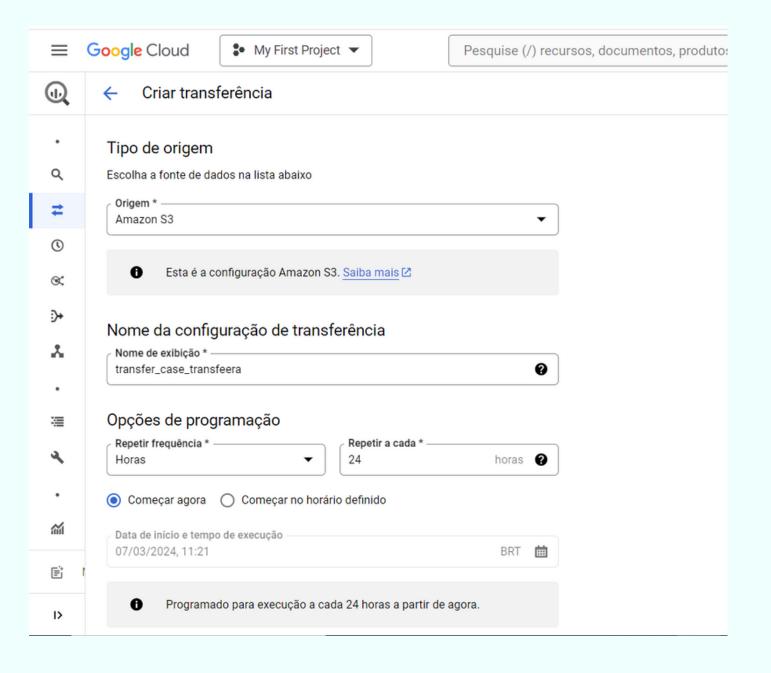
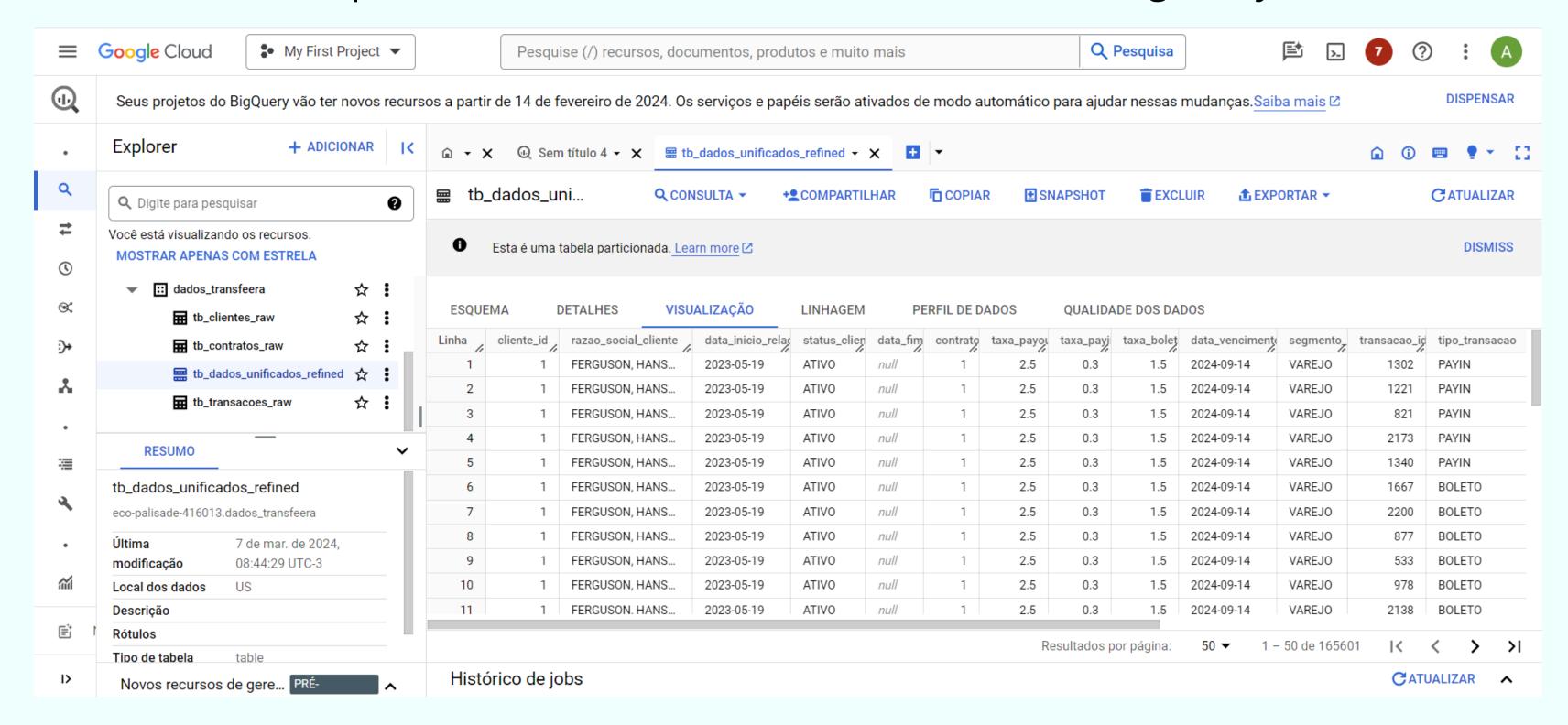


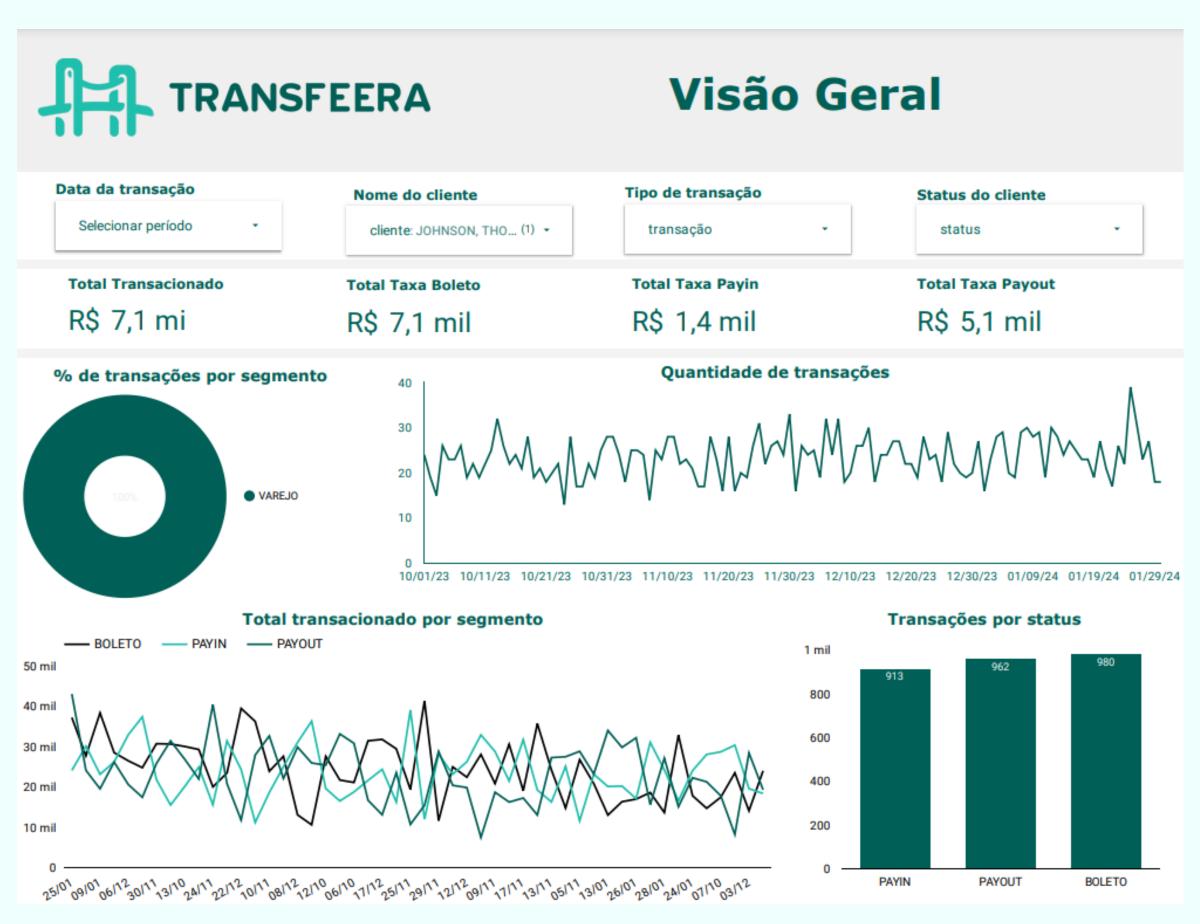


Tabela formada pela união das tabelas raw armazenadas no BigQuery



Visualização de Indicadores





Links para o arquivo:



- Modelo de DAG python
- Dashboard
- Jupyter Notebook desenvolvido para testes locais

https://drive.google.com/drive/folders/1EdLsTsfDTNo5BQUwVNRNwKBzWr5wSYxv?
usp=sharing

https://lookerstudio.google.com/reporting/dcbf63bb-b35c-45b1-b90b-bccb408a651f

OBS: O dashboard pode apresentar lentidão devido aos recursos limitados da versão de teste da GCP

OBRIGADA

Amanda Louise Costa Nascimento

