



Ingestão de dados

Relatório - Atividade 4

Lucas Bonilha
Tiago Galinari
Evander Siqueira
Thiago Pereira

1 Camada de Dados

Com os notebooks de **extração**, **data wrangling** e **modelagem** criamos um pipeline para carregar todos os dados para as camadas **raw**, **trusted** e **analytics**.

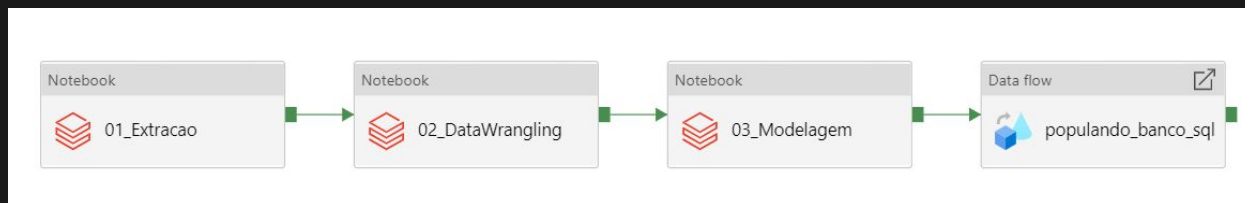
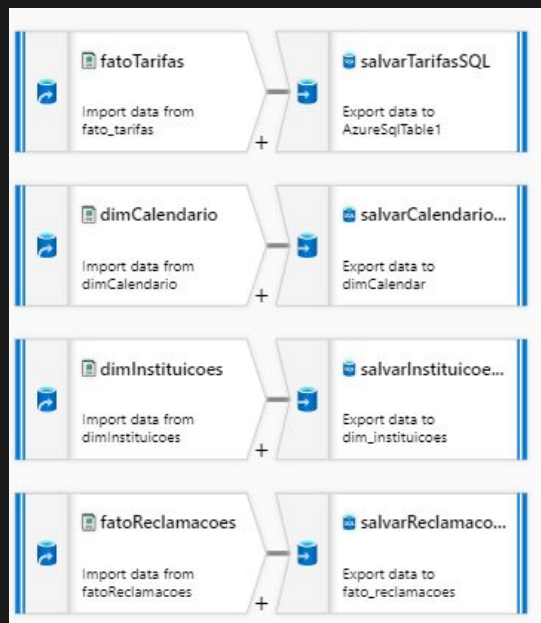


Esse processo foi orquestrado por meio do **Azure Data Factory**. Assim, cada notebook gera uma camada e salva no **Azure Blob Storage**.

Nome	Modificado
<input type="checkbox"/> analytics	
<input type="checkbox"/> raw	
<input type="checkbox"/> trusted	

2 Banco SQL

Ainda no Data Factory, populamos o banco **tarefa4** no Azure SQL com um dataflow.

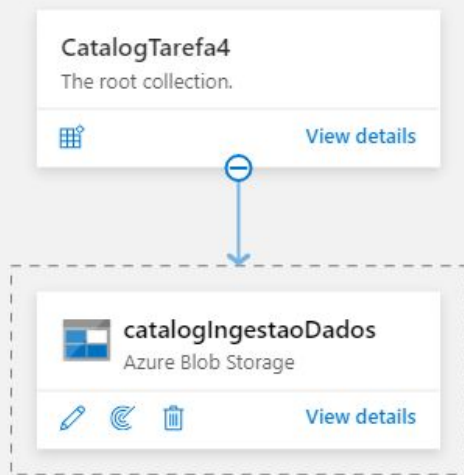


Assim, em uma ordem:

1. Extraímos os dados da API - **raw**
2. Fazemos uma limpeza e tratamento desses dados - **trusted**
3. Modelamos com o SQL - **analytics**
4. E populamos o banco com os dados da Analytics

3 Catálogo

Para catalogar, utilizamos o Microsoft Azure Purview. Nele, mapeamos o blob storage que salvamos as camadas **trusted** e **analytics**.



The screenshot shows the Microsoft Azure Purview catalog interface. The left sidebar displays the 'CatalogTarefa4' collection and a list of related items: Dashboards, Data pipelines, Files, Folders, Reports, Stored procedures, and Tables. The main area shows a list of 13 results, with the first five visible: 'df_reclamacoes.csv', 'fato_reclamacoes.csv', 'fato_tarifas.csv', 'df_lista_inst.csv', and 'dim_calendario.csv'. Each item is a CSV file stored in Azure Blob Storage. The interface includes filters for 'Source type' and 'Instance', and a pagination bar at the bottom showing 'Page 1 of 1'.

Obrigado!