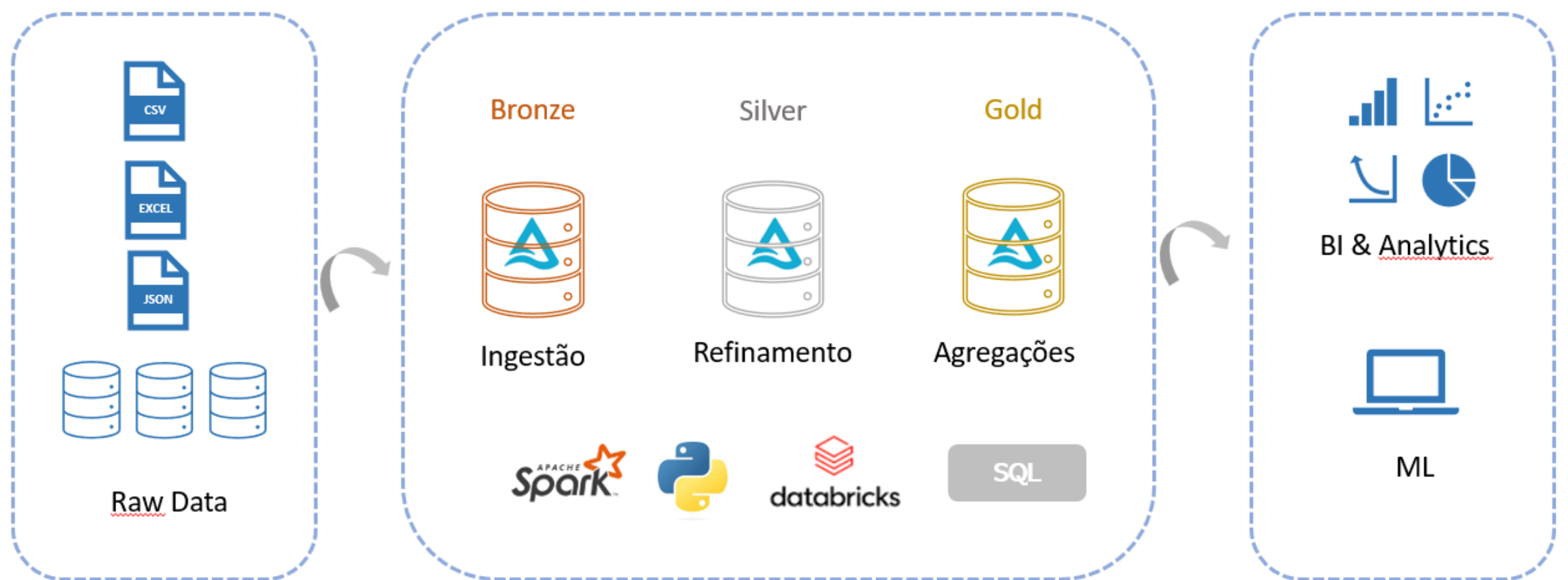


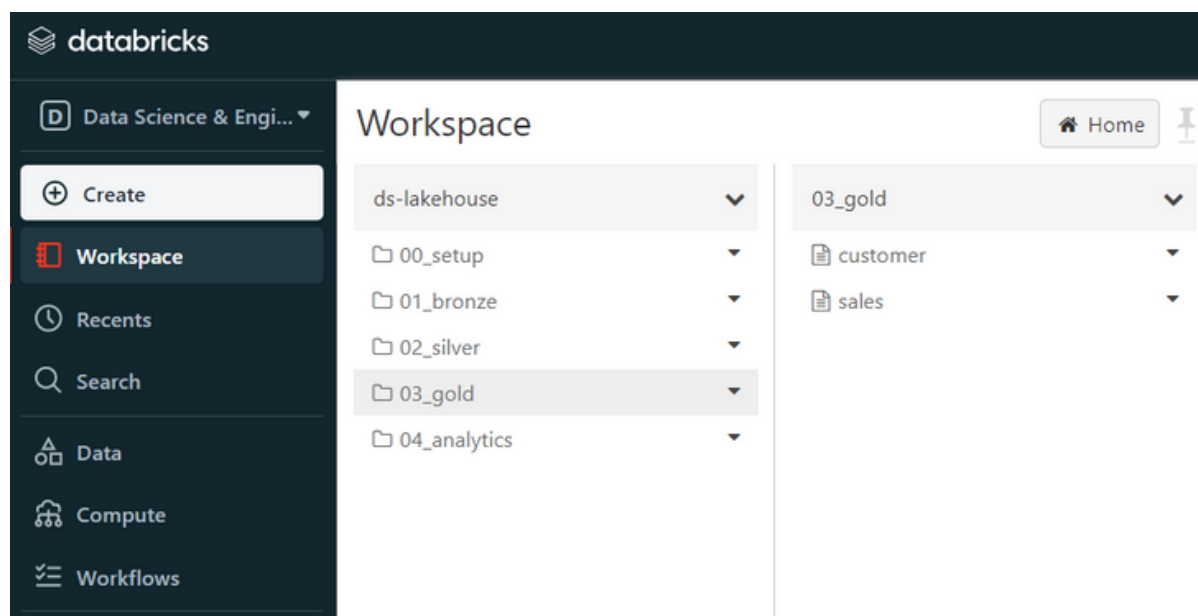
Projeto

Arquitetura Lakehouse

com Databricks

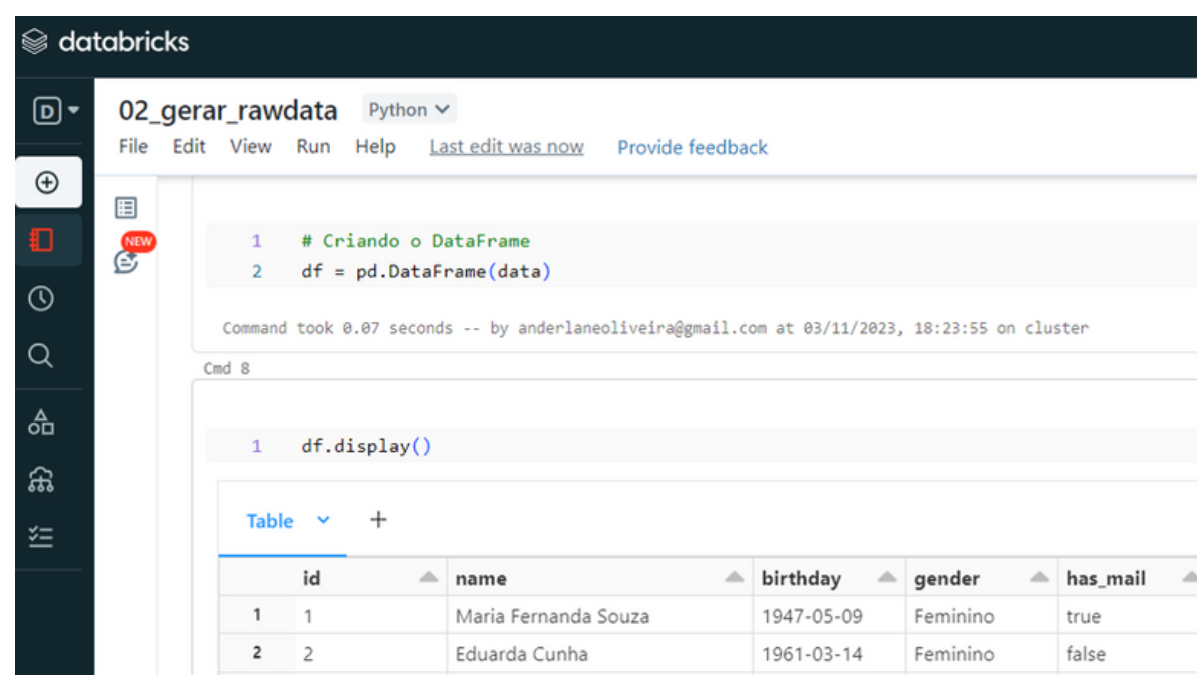


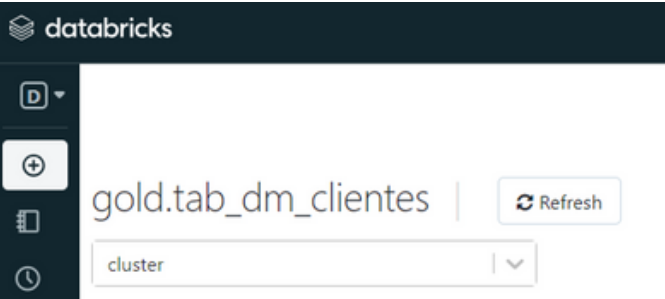
Pipeline de dados para processamento de arquivos em batch



Organização dos processos para desenvolvimento do Lakehouse no Databricks.

Dados fictícios gerados com a biblioteca Python “Faker” e armazenados no DBFS, sistema de arquivos distribuídos, disponível em cluster.



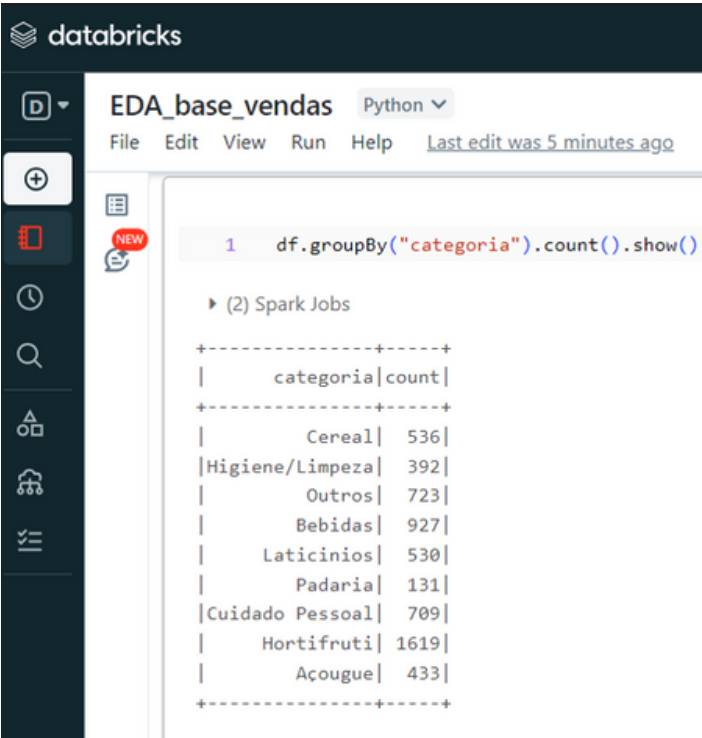


Schema:

	col_name ▲	data_type ▲
1	idCliente	string
2	nome	string
3	idade	bigint
4	genero	string
5	flg_email	int
6	flg_tel	int
7	uf	string

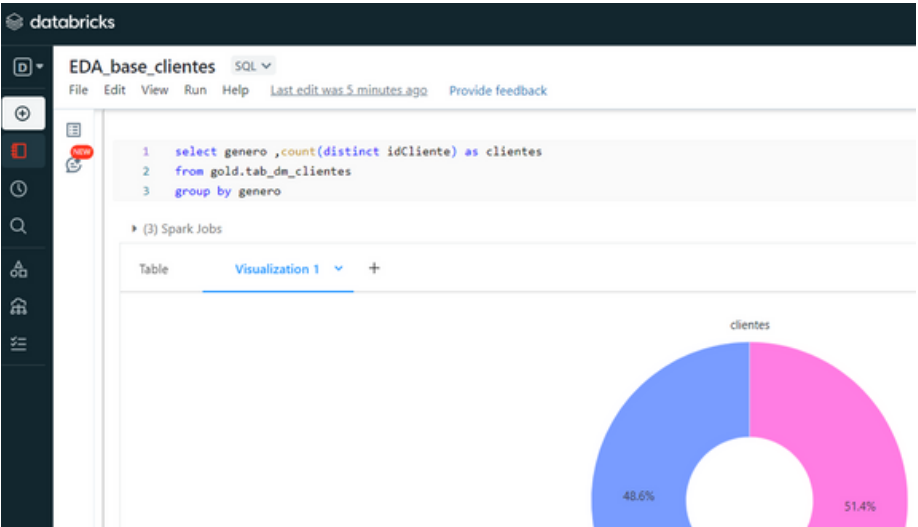
Etapas de Processamento:

- 1- **Camada Bronze:** Ingestão dos dados brutos em estrutura relacional;
- 2- **Camada Silver:** Transformação, Limpeza e Refinamento dos dados;
- 3- **Camada Gold:** Agregações e Carga para uso dos dados em consultas.



Dados disponíveis para consultas em linguagens Python, R, SQL ou Scala.

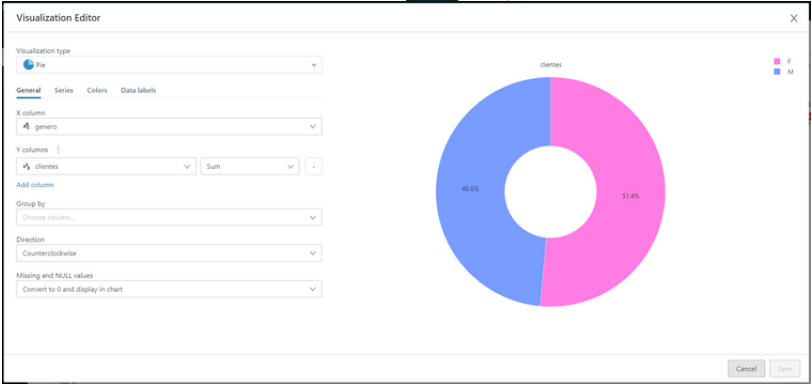
As consultas realizadas podem ser visualizadas em formato de tabela e adição de diversos tipos de gráficos.



(3) Spark Jobs

Table Visualization 1

	genero	clientes
1	F	857
2	M	809



The image features a white background with the word "Thanks!" centered in a bold, black, sans-serif font. In each of the four corners, there is a cluster of overlapping, semi-transparent geometric shapes, primarily squares and rectangles, in various shades of blue and teal. These shapes are arranged in a way that they appear to be floating or layered, creating a modern, abstract design.

Thanks!