



Microsoft Azure



Microsoft

Azure
Academy

e.Seth Cloud

Proposta de Solução de Negócio

CASE 4: Uso dos dados de Extrato

EQUIPE 4



BRDESCO – CASE 4

HISTÓRICO DO DOCUMENTO				
Versão	Elaborado por	Data	Revisado por	Data
1	Flavia Nitto	11/03/2023	Luiz Galante	12/03/2023

COMPOSIÇÃO EQUIPE
Douglas Graciano Littig
Flavia Narumi Nitto
Laura Aparecida Mariano da Silva
Lilia Bakker
Luiz Gustavo Tanaami Galante
William Kliemann Junior



1 Introdução

Esse documento tem como objetivo difundir a solução proposta pela equipe 04, abordando o Case 4 apresentado no Open Finance Hackathon, realizado pela e.Seth Cloud, Azure Academy e Bradesco com patrocínio da Microsoft.

2 Detalhamento do Case

O case 4 porta um problema de negócio, no qual o Bradesco busca por meio das tecnologias concernentes aos ambientes de dados, prover insights e alcançar leads de qualidade.

É demandado que uma solução para recategorizar os extratos 1ª linha e 2ª linha, de maneira geral a primeira linha não traz tantos ganhos, pois, tem fácil categorização, devido a ter 18 classificações específicas, informadas pelo BACEN (Banco Central do Brasil). Então, o desafio do case está na segunda linha, onde cada entidade bancária pode, conforme deliberação própria, alterar e/ou mudar qualquer informação descrita.

3 Abordagem da Escolha do Case

A Equipe 4 observou nesse desafio, uma grande oportunidade de negócios para o Bradesco, pois, fica evidente a gama de serviços que podem alcançar os leads corretos e atingindo metas de maneira mais assertiva e assim gerar resultados expressivos em campanhas personalizadas.

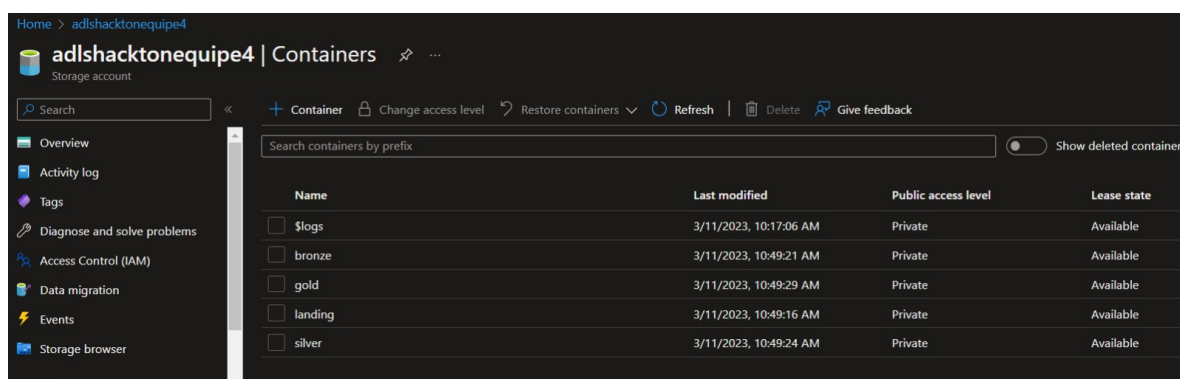


4 Passo a passo do ETL

4.1 Criação da infraestrutura

4.2 Acessar o datalake do banco

4.3 Copiar para o lake privado do grupo - ingestão de dados



Name	Last modified	Public access level	Lease state
<input type="checkbox"/> \$logs	3/11/2023, 10:17:06 AM	Private	Available
<input type="checkbox"/> bronze	3/11/2023, 10:49:21 AM	Private	Available
<input type="checkbox"/> gold	3/11/2023, 10:49:29 AM	Private	Available
<input type="checkbox"/> landing	3/11/2023, 10:49:16 AM	Private	Available
<input type="checkbox"/> silver	3/11/2023, 10:49:24 AM	Private	Available

4.4 Pipeline da Azure - ETL completo

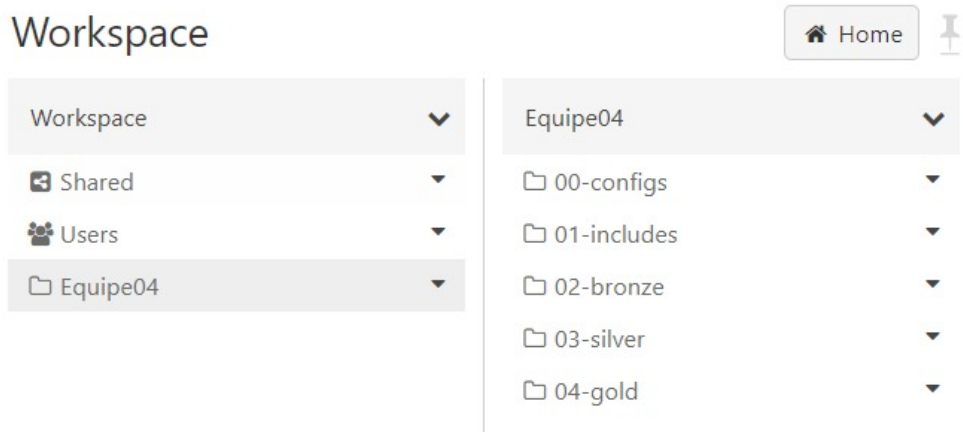
- Copy data extraíndo o dado do lake *hackerhub* e trazendo para a nossa primeira camada de ingestão
- Notebook de ingestão levando o dado da camada *landing* para a *bronze* (lakehouse)
- Notebook de tratamento dos dados brutos (*bronze*) para a camada *silver*
- Notebook de análises levando o resultado obtido em cima da camada *silver*, até a camada *gold*.



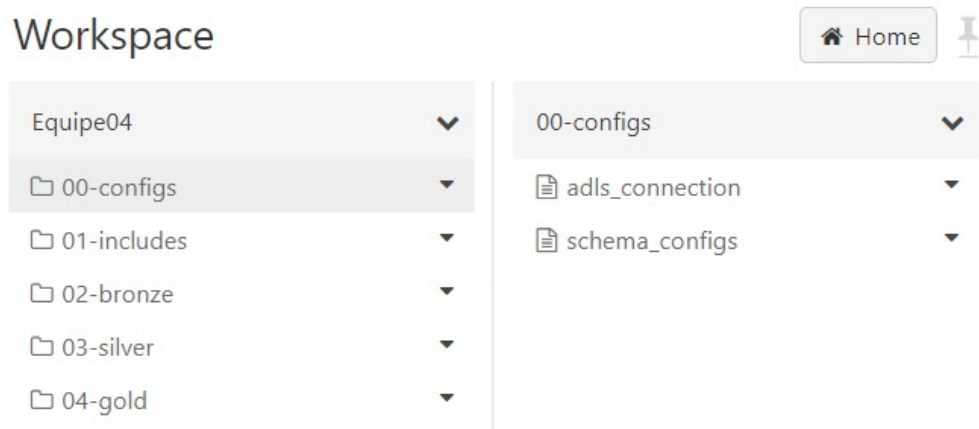


4.5 Tratamento de dados

- Organização das pastas no workspace do Databricks



- **00-configs:** contém os arquivos pertinentes à configuração do ambiente, notebook para montagem do ADLS, path de pastas, etc.





- **02-bronze:** essa pasta contém os notebooks que fazem a ingestão da camada landing para a camada bronze.

Workspace Home

Equipe04	02-bronze
00-configs	accounts_transactions-landing-to-...
01-includes	consents-landing-to-bronze
02-bronze	
03-silver	
04-gold	

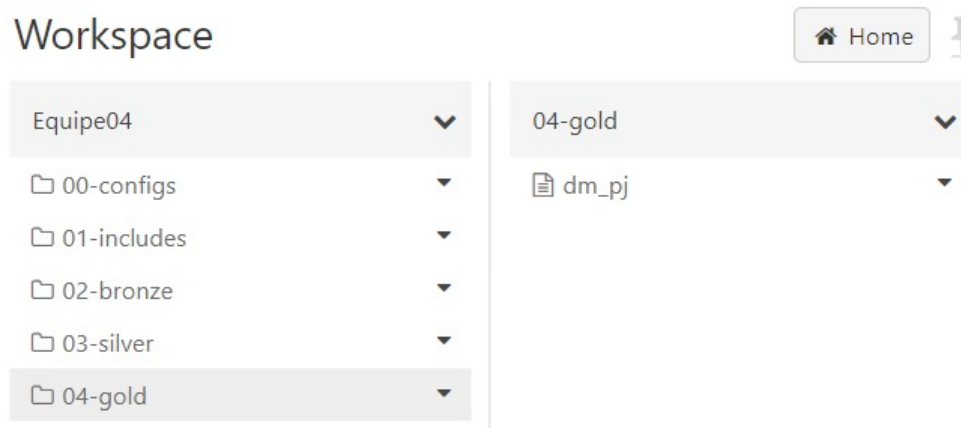
- **03-silver:** essa pasta contém os notebooks que foram utilizados para realizar a análise exploratória dos dados e também o notebook que faz o tratamento das tabelas bronzes e grava na camada silver.

Workspace Home

Equipe04	03-silver
00-configs	Analises
01-includes	transformacao_silver
02-bronze	
03-silver	
04-gold	



- **04-gold:** essa pasta contém o notebook com as análises efetuadas utilizando SQL para identificação de padrões dentro da segunda linha do extrato, e também grava o resultado de cada análise em um arquivo CSV.





5 Filtragem de leads qualificados

Após os dados serem ingeridos e tratados, foi feita uma análise dos dados e encontrado alguns padrões de palavras chave, como por exemplo “seguro”. Além disso, ao aplicar uma busca em toda a base, encontramos ocorrências do mesmo padrão em outras transações de outros bancos. Possibilitando a criação de uma *word list* onde poderemos treinar o Machine Learning.

transactionId	transactionDate	transactionName	creditDebitType	amount	brandName
e1ab0d92e9622549307185855f7ada61edc17bb94b00005a5f778689c2368212	2022-09-18	SxGhRy wTzhxMPRxSz 0/0	DEBITO	-217.13	Banco 3
cc1fac28a7f5ca1eb574e8c0c6d732667079d9ac13ee756f0af0e38e4c587da	2022-11-18	SxGhRy wTzhxMPRxSz 0/0	DEBITO	-217.13	Banco 3
f0f8ac6c9c143827177e02dc5de44e69b2a6b216ba9538262bedddb0d8024f7a	2022-11-18	SxGhRy wTzhxMPRxSz 0/0	DEBITO	-217.13	Banco 3
b4370a61cc50ebd820ee3d9358bba099ff59e16f32bb73911cd070317fc88df0	2022-08-18	SxGhRy wTzhxMPRxSz 0/0	DEBITO	-217.13	Banco 3
83a676cc817410c73f22c6992c9cd53a2efafeab4a3a1f2ba3993dc22a4cc770	2022-10-19	SxGhRy wTzhxMPRxSz 0/0	DEBITO	-217.13	Banco 3
fa9403bf092a20c909a172271aec5c4ff7ef51abf45d3c744d525abef6c4374d	2022-07-21	SxGhRy wTzhxMPRxSz 0/0	DEBITO	-217.13	Banco 3

transactionId	transactionDate	transactionName	creditDebitType	amount	brandName
99da13934ef50ab9aa778cdac8b8d9a008995f6f68dbae18	2022-07-21	zPLwCzCzy zhT CyNTzMzX xMPRxSzRwzL	DEBITO	-324.81	Banco 39
iac49c8d06219fed1caa0ea56ccab826d0a0f8437ebca263	2022-10-14	zPLwCzCzy zhT CyNTzMzX xMPRxSzRwzL	DEBITO	-28312.98	Banco 39
af980ac5208192a94073748b76615ffb2d5d57f17a4a9688	2022-10-09	zPLwCzCzy zhT CyNTzMzX xMPRxSzRwzL	DEBITO	-389.4	Banco 39
b329a94a0d3c2027fc263200ebe8a8f247344776a8ac488	2022-10-06	zPLwCzCzy zhT CyNTzMzX xMPRxSzRwzL	DEBITO	-974.83	Banco 39
f8325a57ba3d6b40d38ab81bec8bed8daa940eb8e9598d7b	2022-08-21	zPLwCzCzy zhT CyNTzMzX xMPRxSzRwzL	DEBITO	-333.8	Banco 39
4b235f5f94b00104e8b7acce2013dc8e7d5955fbc947592	2022-10-26	zPLwCzCzy zhT CyNTzMzX xMPRxSzRwzL	DEBITO	-1444.43	Banco 39