

SOUL CODE



Google Cloud

MIGRAÇÃO E INTEGRAÇÃO DE BASES E DADOS

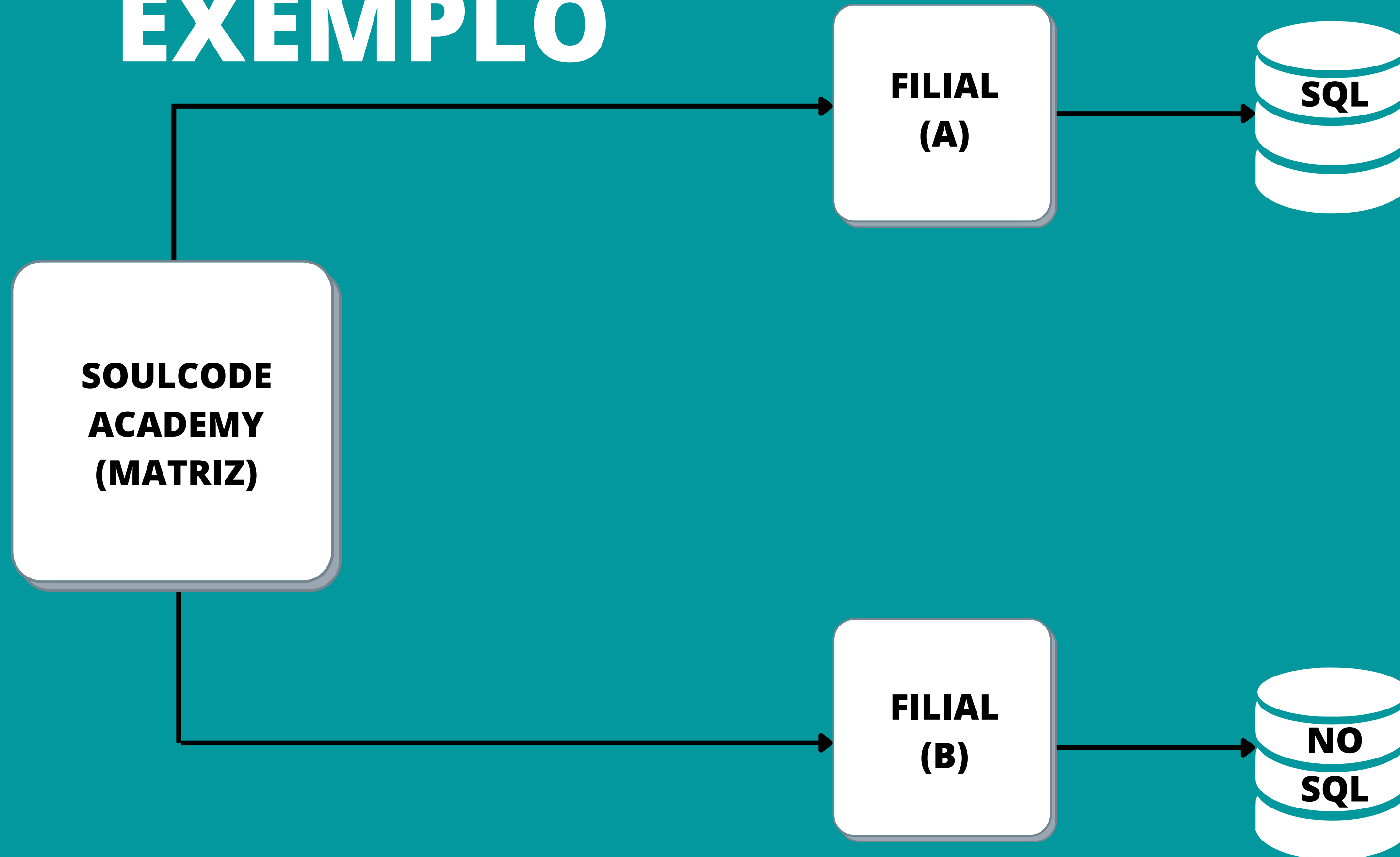
ATIVIDADE 20

Apresentada por Edmar, Estudante de Engenharia de dados

PROBLEMA

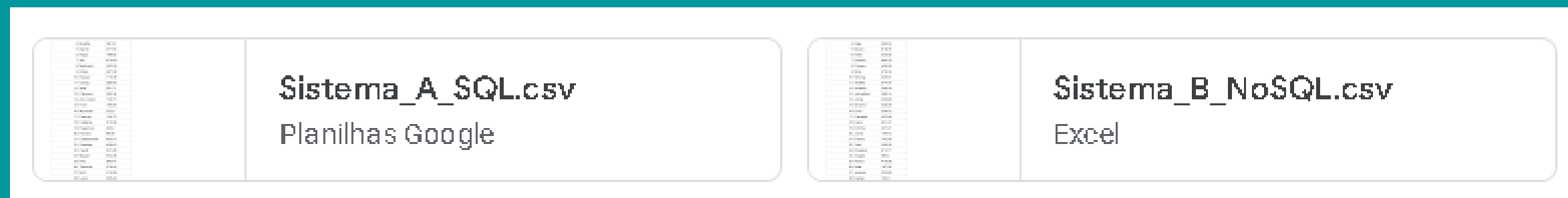
A SoulCode Academy utiliza dois sistemas distintos para cadastro de vendas, um em sua matriz e outro em suas filiais. O sistema da matriz utiliza um db noSQL enquanto os das filiais ainda utilizam um SQL. As tabelas a seguir são amostras de dados dos dois tipos de db enviadas para auxiliar seu desenvolvimento. Crie uma db em SQL e uma noSQL, e adicione essas amostras para criar um ambiente simulando o problema real. A seguir crie um código em Python que leia a db SQL, corrija, padronize e insira os dados na db noSQL.

EXEMPLO



DADOS

- Foi dado 2 arquivos da filiais
- Cada arquivo contém as colunas nota_fiscal, vendedor e total
- Com 2000 linhas e 3 colunas

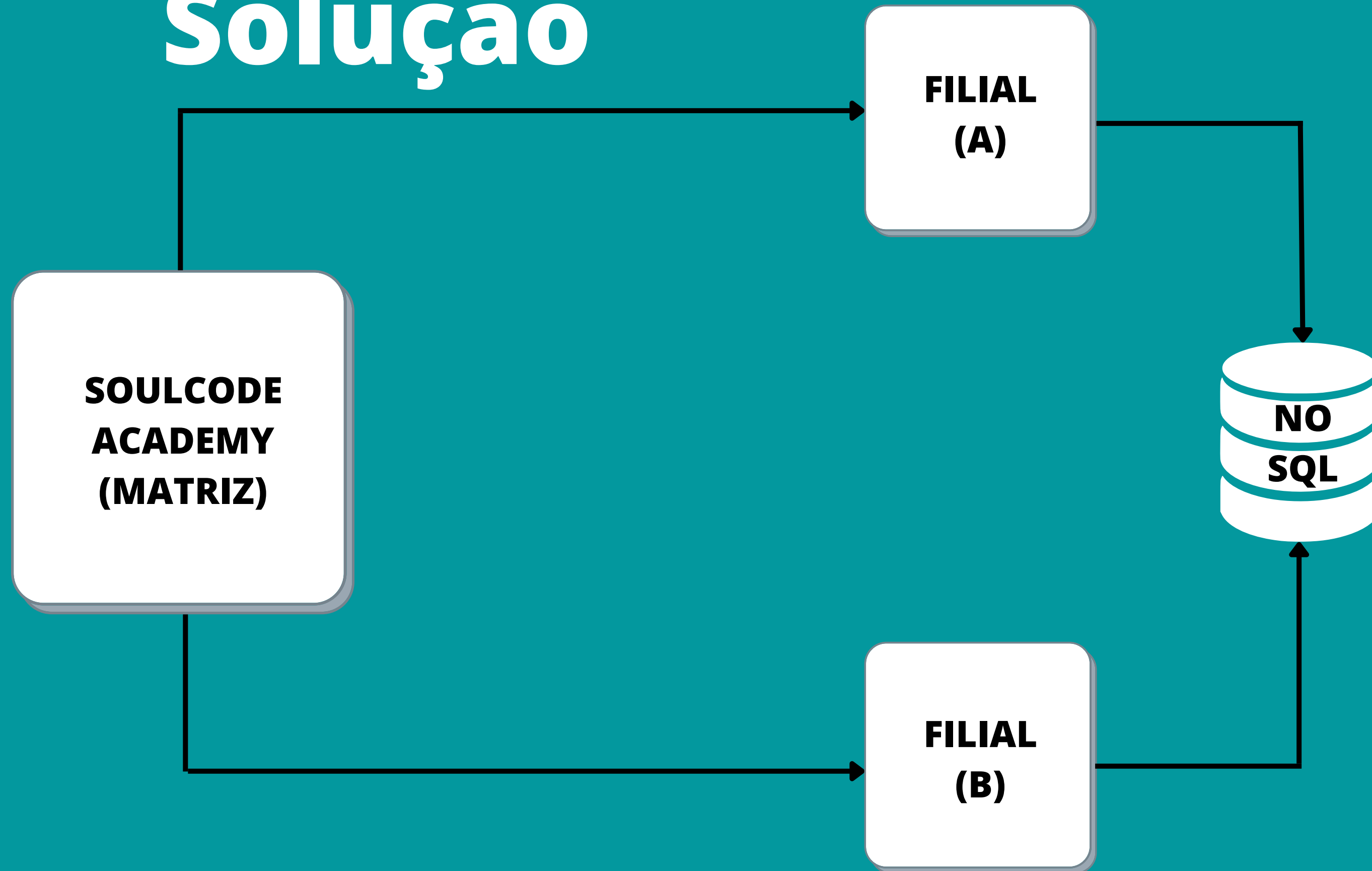


	A	B	C	D
1	nota_fiscal,vendedor,total			
2	1,Clerkclaude,260.85			
3	2,Karly,217.45			
4	3,Xerxes,462.32			
5	4,Kaia,290.52			
6	5,Myron,518.35			
7	6,Eddy,320.92			
8	7,Suellen,464.28			
9	8,Doreen,406.99			
10	9,Mira,579.04			
11	10,Waring,225.81			
12	11,Ardelia,574.07			
13	12,Ariadne,396.09			
14	13,Johnathon,369.16			
15	14,Joline,349.68			
16	15,Brianne,326.26			
17	16,Elvin,336.53			
18	17,Demetre,452.96			
19	18,Karie,251.07			
20	19,Dimitry,447.47			
21	20,Julina,146.43			
22	21,Felicia,142.56			
23	22,Vikki,259.26			
24	23,Rosabel,512.77			
25	24,Angela,584.2			
26	25,Noemi,416.58			
27	26,Neal,187.92			
28	27,Jereme,323.85			
29	28,Darsie,162.0			
30	29,Ashbey,153.29			
31	30,Andrey,348.36			
32	31,Diena,154.44			
33	32,Brook,551.05			
34	33,Wilek,358.25			
35	34,Dorey,161.6			
28	27,Britt,315.99			
29	28,Lanni,305.48			
30	29,Collin,309.09			
31	30,Edwin,281.83			

SOLUÇÃO

Realizar a leitura do banco de dados da filial,
corrigir, padronizar e inserir os dados no
banco de dado da matriz.

Solução



AGENDA

01

- AVALIAÇÃO
20/05/2022

02

- FERRAMENTAS
- DIVISÃO DO
TRABALHO

03

- RESOLUÇÃO

04

- VERIFICAÇÃO
- RESULTADO
23/05/2022

FERRAMENTAS

- **VSCODE** - Para o desenvolvimento dos códigos, normalização e inserção dos dados.
- **GOOGLE CLOUD** - Para criação dos bancos na nuvem.
- **MYSQL WORKBENCH** - Visualização e criação do banco
- **MONGO ATLAS** - Para visualização e monitoramento do banco.

BANCOS DE DADOS

The screenshot displays the Google Cloud Platform console interface. On the left, the 'Cluster0' page shows a SQL editor with the following code:

```
-- Cria o Banco de Dados
CREATE DATABASE filialA;

-- Seleciona o banco de dados
USE filialA;

-- cria a tabela vendas
create table if not exists vendas(
    nota_fiscal integer auto_increment,
    nome_vendedor varchar(30) not null,
    total float not null,
    constraint dados_pkey primary key (nota_fiscal));
```

On the right, the 'Databases' page for instance 'ativ18' (MySQL 8.0) is shown. A red arrow points to the 'filialA' database in the list.

Name ↑	Collation	Character set	Type
ativ18	utf8_general_ci	utf8	User
filialA	utf8mb4_0900_ai_ci	utf8mb4	User
guestbook	utf8mb4_0900_ai_ci	utf8mb4	User
information_schema	utf8_general_ci	utf8	System
mydatabase	utf8mb4_0900_ai_ci	utf8mb4	User
mysql	utf8_general_ci	utf8	System
performance_schema	utf8mb4_0900_ai_ci	utf8mb4	System
sys	utf8mb4_0900_ai_ci	utf8mb4	System

- Foram criados os bancos na nuvem do google cloud e usando o Mysql Workbench e MongoAtlas para visualização dos dados

BANCOS DE DADOS

Pelo VSCODE foi feito o tratamento, conexão e inserção dos dados de cada banco.

```
#Leitura de dados e maximizando exibição
df = pd.read_csv("Sistema_A_SQL.csv", sep=",")
pd.options.display.max_rows = 1000
pd.options.display.max_columns = 3
#df
✓ 0.1s

#Excluindo a coluna nota_fiscal do DataFrame para inserção nos databases
df.drop(['nota_fiscal'], axis=1, inplace=True)
✓ 0.1s

#Exibindo o resultado
#df
df.dtypes
df.head()
✓ 0.1s
```

```
#Conector para o banco SQL
import mysql.connector

mydb = mysql.connector.connect(
    host="35.222.110.4",
    user="root",
    password="Eddiesp16adm",
    database="filialA"
)

print("Conectado!", mydb)
```

BANCOS DE DADOS

Pelo VSCODE foi feito o tratamento, conexão e inserção dos dados de cada banco.

```
#Insere os dados do DataFrame filia_A_SQL no banco de dados SQL -- MySQL ...
mycursor = mydb.cursor()
for index, row in df.iterrows():
    query = f"INSERT INTO vendas (nome_vendedor, total) VALUES ('{row['vendedor']}', '{int(row['total'])}')"
    mycursor.execute(query)
```

```
# Efetiva a inserção e exibe quantas inserções foram feitas
mydb.commit()

print(mycursor.rowcount, "inserção concluída.")
```

1 was inserted.

BANCOS DE DADOS

Foi feito a exclusão dos valores nulos de ambos os dados ficando com ao todo 1929 linhas de 2000.

```
#Excluindo a coluna nota_fiscal do DataFrame para inserção nos databases  
df2.drop(['nota_fiscal'], axis=1, inplace=True)
```

✓ 0.3s

```
#Excluindo as linhas que contem campos vazios para inserção no database  
df2 = df2.dropna()  
df = df.dropna()
```

✓ 0.1s

RESOLUÇÃO

Depois de feito a conexão usando o pymongo, foi convertido os 2 dataframes em dicionários e inseridos no mongo.

```
#conector para mongodb
import pymongo

myclient = pymongo.MongoClient("mongodb+srv://root:Eddiesp16@cluster0.4u1yb.mongodb.net/filialb.filialb")

mydb = myclient["filialb"]

mycol = mydb["filialb"]
```

✓ 5.9s

```
#inserindo os dados no banco mongo convertendo para dicionarios
data = df2.to_dict(orient = "records")
db = myclient["filialb"]
db.filialb.insert_many(data)
```

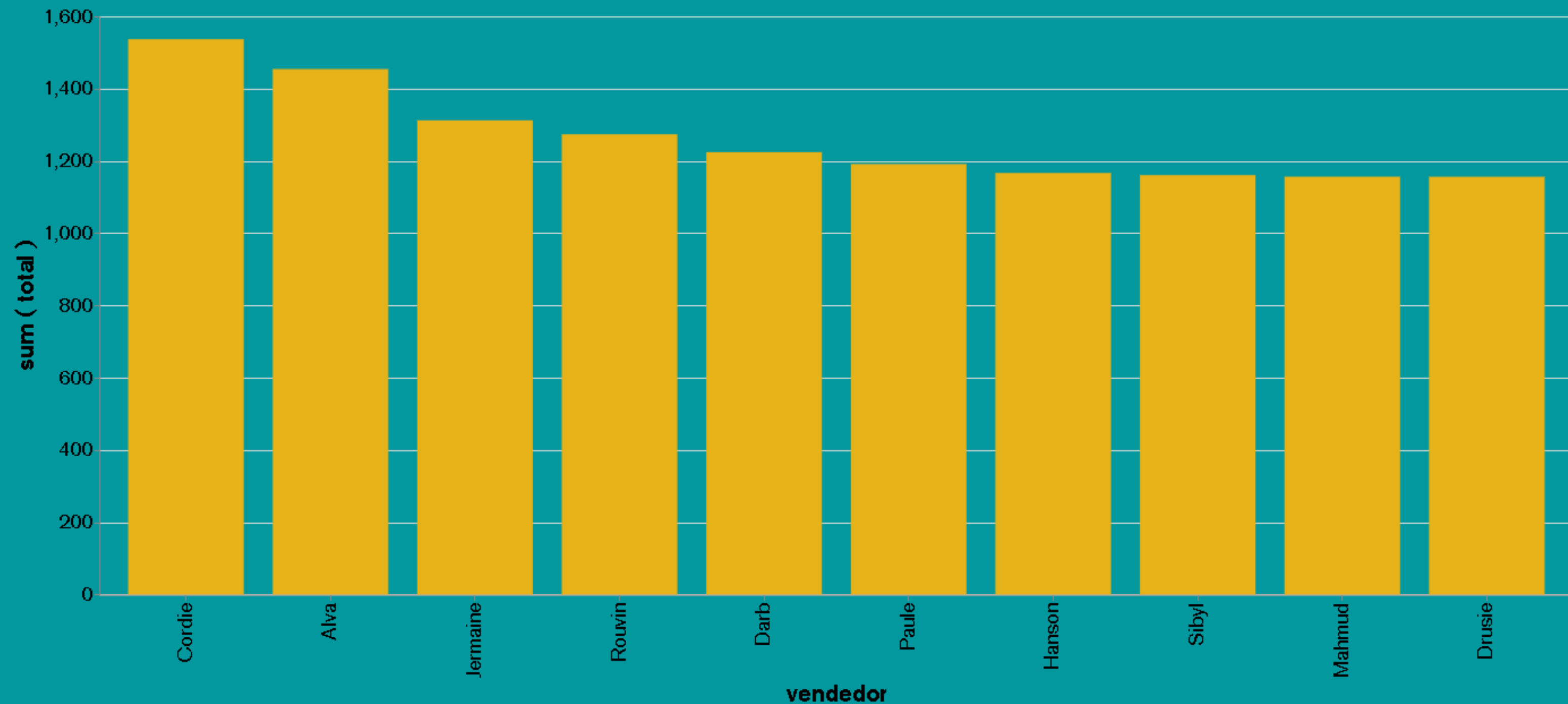
✓ 1.3s

pymongo.results.InsertManyResult at 0x1fde68d3440>

```
#inserindo os dados do mysql no banco mongo convertendo para dicionarios
data = df.to_dict(orient = "records")
db = myclient["filialb"]
db.filialb.insert_many(data)
```

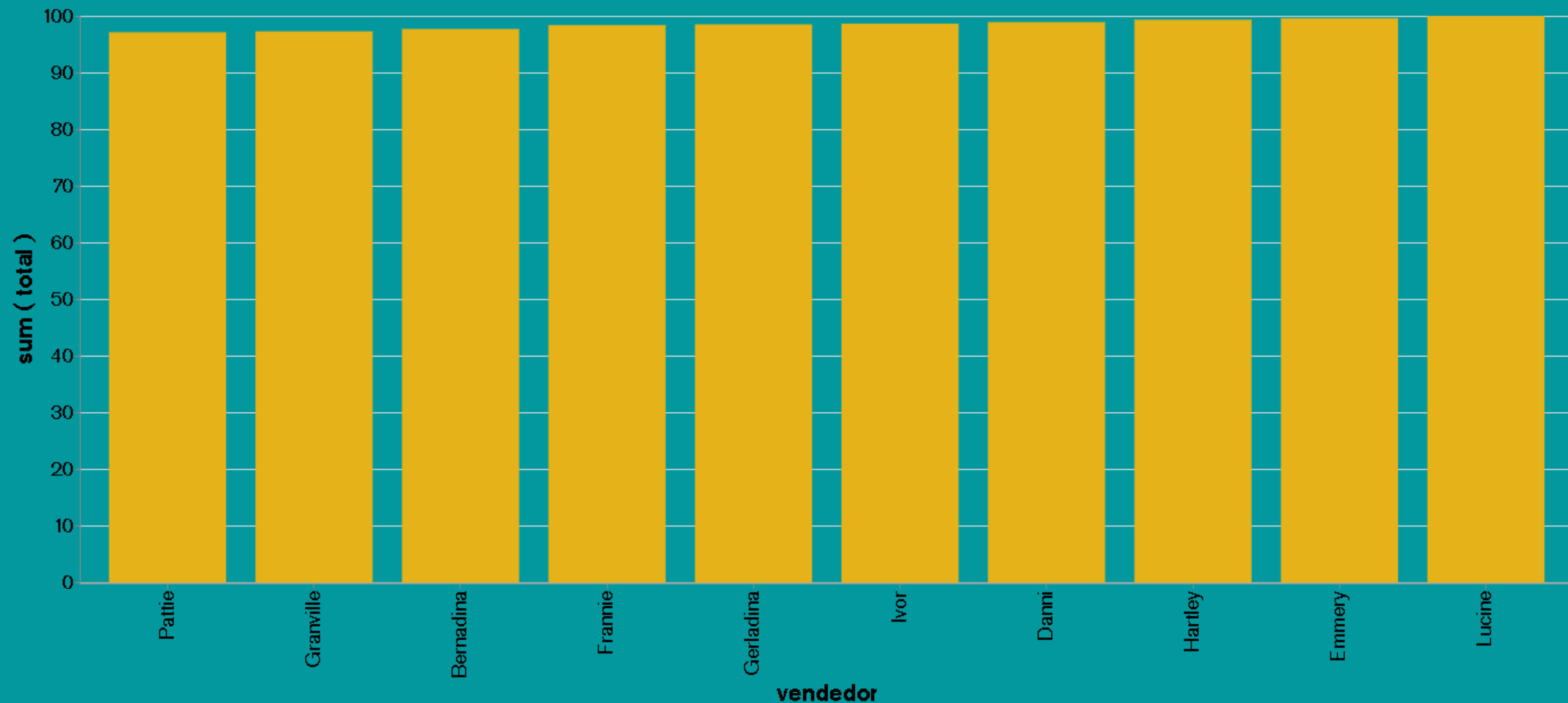
✓ 0.3s

INSIGHTS



OS 10 VENDEDORES COM MAIS VENDAS

INSIGHTS



OS 10 VENDEDORES COM MENOS VENDAS

MELHORIAS

- Incluir outros banco no mesmo processo como cassandra.
- Criação de uma interface gráfica com tkinter ou Pysimplegui com instalador de pacotes miconda e criar o executável com pyinstaller para rodar como um aplicativo de início para facilitar o processo de transferência de dados e deixar mais acessível aos outros colaboradores

FONTES

- w3schools.com
- <https://stackoverflow.com/>
- <https://github.com/>
- Documentação do Google Cloud

OBRIGADO!