

4. Extração de Dados - Usuário

O objetivo deste documento, é facilitar a extração dos dados pelo usuário

Importação de bibliotecas

```
In [1]: import os
import shutil
from pyspark.sql import SparkSession
from pyspark.sql.functions import from_unixtime, col, to_timestamp, coalesce
from pyspark.sql.types import StringType, IntegerType, LongType, DecimalType, DateType
```

Variáveis do projeto

```
In [2]: #Diretorio dos arquivos csv
v_diretorio_csv='/usr/local/spark/csv/'

#Diretorio de export do arquivo de flatfile
v_diretorio_export='/home/jovyan/work/export'

#Nome do arquivo do export
v_nome_arquivo='movimento_flat.csv'

#Variaveis de conexao com postgres
v_caminho_jar_postgres='/home/jovyan/work/jars/postgresql-9.4.1207.jar'
v_url_jdbc='jdbc:postgresql://postgres/projeto'
v_user_jdbc='airflow'
v_pass_jdbc='airflow'
```

Criando sessao e contexto

```
In [3]: spark = (SparkSession
    .builder
    .master('local')
    .appName('load-postgres')
    # Add postgres jar
    .config('spark.driver.extraClassPath', v_caminho_jar_postgres)
    .getOrCreate())
sc = spark.sparkContext
```

```
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/usr/local/spark-3.1.2-bin-hadoop3.2/jars/spark-unsafe_2.12-3.1.2.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
22/12/23 09:30:02 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
```

Geração do arquivo Flat

O objetivo do exercício, é exercitar o processamento no Spark. A extração dos dados e relacionamentos entre os dados, serão processados no **Spark**.

Por essa razão, os dados não serão relacionados e tratados no **postgres**.

In [4]:

```
#Carregando dados no Dataframe
df_associado_tgt = (
    spark.read
    .format('jdbc')
    .option('url', v_url_jdbc)
    .option('dbtable', 'target.associado')
    .option('user', v_user_jdbc)
    .option('password', v_pass_jdbc)
    .load()
)

#Criando view do dataframe
df_associado_tgt.createOrReplaceTempView('associado')
```

In [5]:

```
#Carregando dados no Dataframe
df_conta_tgt = (
    spark.read
    .format('jdbc')
    .option('url', v_url_jdbc)
    .option('dbtable', 'target.conta')
    .option('user', v_user_jdbc)
    .option('password', v_pass_jdbc)
    .load()
)

#Criando view do dataframe
df_conta_tgt.createOrReplaceTempView('conta')
```

In [6]:

```
#Carregando dados no Dataframe
df_cartao_tgt = (
    spark.read
    .format('jdbc')
    .option('url', v_url_jdbc)
    .option('dbtable', 'target.cartao')
    .option('user', v_user_jdbc)
    .option('password', v_pass_jdbc)
    .load()
)

#Criando view do dataframe
df_cartao_tgt.createOrReplaceTempView('cartao')
```

In [7]:

```
#Carregando dados no Dataframe
df_movimento_tgt = (
    spark.read
    .format('jdbc')
    .option('url', v_url_jdbc)
    .option('dbtable', 'target.movimento')
    .option('user', v_user_jdbc)
    .option('password', v_pass_jdbc)
    .load()
)

#Criando view do dataframe
df_movimento_tgt.createOrReplaceTempView('movimento')
```

Obs: 1 - Somente será trabalhado com associados válidos. 2 - Na estrutura de dados gerados, não foi repassado a coluna de data da criação do cartão, temos somente a data de criação da conta, por essa razão, será utilizada a data de criação da conta, como data de criação do cartão.

```
In [8]: df_flat_file=spark.sql('''
    select
        ass.nome as nome_associado,
        ass.sobrenome as sobrenome_associado,
        ass.idade as idade_associado,
        mov.vlr_transacao as vlr_transacao_movimento,
        mov.des_transacao as des_transacao_movimento,
        mov.data_movimento as data_movimento,
        car.num_cartao as numero_cartao,
        car.nom_impreso as nome_impreso_cartao,
        cco.data_criacao as data_criacao_cartao,
        cco.tipo as tipo_conta,
        cco.data_criacao as data_criacao_conta

    from associado ass

    left join cartao car
    on car.id_associado=ass.id

    left join movimento mov
    on mov.id_cartao=car.id

    left join conta cco
    on cco.id_associado=ass.id

    where ass.id>0

''')
```

```
In [9]: df_flat_file.show()
```

[Stage 5:=====> (189 + 1) / 200]

nome_associado	sobrenome_associado	idade_associado	vlr_transacao_movimento	des_transacao_movimento	data_movimento	numero_cartao	nome_impreso_cartao	data_criacao_cartao	tipo_conta	data_criacao_conta
Valentina	Fogaça	23	159.36	Supermer	2018-12-13	4250002301488069	VALENTINA FOGAÇA	2018-08-14	Conta Corrente	2018-08-14
Valentina	Fogaça	23	241.92	Restaur	2020-06-06	4250002301488069	VALENTINA FOGAÇA	2018-08-14	Conta Corrente	2018-08-14
Valentina	Fogaça	23	154.46	R	2020-06-16	4250002301488069	VALENTINA FOGAÇA	2018-08-14	Conta Corrente	2018-08-14
Valentina	Fogaça	23	188.04	Restaur	2020-09-03	4250002301488069	VALENTINA FOGAÇA	2018-08-14	Conta Corrente	2018-08-14
Valentina	Fogaça	23	196.40	Posto combust	2019-08-24	4250002301488069	VALENTINA FOGAÇA	2018-08-14	Conta Corrente	2018-08-14
Valentina	Fogaça	23	59.40	Restaur	2020-03-26	4250002301488069	VALENTINA FOGAÇA	2018-08-14	Conta Corrente	2018-08-14
Valentina	Fogaça	23	225.65	Restaur	2020-12-02	4250002301488069	VALENTINA FOGAÇA	2018-08-14	Conta Corrente	2018-08-14
Valentina	Fogaça	23	2.88	Pet	2018-12-20	4250002301488069	VALENTINA FOGAÇA	2018-08-14	Conta Corrente	2018-08-14
Valentina	Fogaça	23	41.43	Posto combust						

ivel	2020-02-13 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	271.61 R
oupa	2020-04-23 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	63.81 Restaur
ante	2019-02-02 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	242.82 Posto combust
ivel	2020-12-24 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	66.45 Supermer
cado	2020-02-14 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	293.68 Farm
acia	2021-05-25 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	134.12 Restaur
ante	2021-06-07 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	31.12 R
oupa	2018-12-10 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	378.84 R
oupa	2019-11-11 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	122.03 Farm
acia	2021-09-04 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	178.75 Pet
shop	2018-12-11 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
	Valentina Fogaça	23	252.25 R
oupa	2019-11-25 4250002301488069	VALENTINA FOGAÇA	2018-08-14 Conta Corrente
+-----+-----+-----+-----+			
---+-----+-----+-----+-----+			
-----+			
only showing top 20 rows			

In [10]:

In [11]:

```
[Stage 13:> (0 + 1) / 1]
Arquivo flat gerado: /home/jovyan/work/export/movimento_flat.csv
```