RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

ANÁLISE DE DADOS TRANSFORMANDO ESTRUTURA DE SISTEMA DE ANÁLISE ESTATÍSTICA (SAS) EM PYSPARK

DATA ANALYSIS TRANSFORMING A STATISTICAL ANALYSIS SYSTEM (SAS) STRUCTURE INTO PYSPARK

ANÁLISIS DE DATOS QUE TRANSFORMA EL MARCO DEL SISTEMA DE ANÁLISIS ESTADÍSTICO (SAS) EN PYSPARK

Tiago Veiga[1], Fabiana Florian[2]

e5126033
https://doi.org/10.47820/recima21.v5i12.6033

PUBLICADO: 12/2024

RESUMO
Este trabalho tem o objetivo de fazer a migração do código de uma linguagem de programação de Sistema de Análise Estatística (SAS) para a linguagem de programação PySpark em uma empresa. Com essa ferramenta é possível processar uma grande quantidade de dados, aproveitando as vantagens oferecidas pela computação distribuída e pela flexibilidade do ecossistema Python. Foi realizada pesquisa bibliográfica e a transcrição do *script* desenvolvido em SAS para PySpark em uma empresa de tecnologia do Município de Araraquara - SP. Conclui-se que a realização da lógica de programação em PySpark permitiu trazer para o ambiente corporativo uma ferramenta que possui diversas bibliotecas para análises de dados, com agilidade no processamento, confiabilidade na manipulação, que tem como foco o processamento de volumes grandiosos de dados, além de facilitar a elaboração de estudos e a entrega de informações.

PALAVRAS-CHAVE: Análises. Desenvolvimento. Dados. PySpark. Python. SAS.

ABSTRACT
This work aims to migrate the code from a Statistical Analysis System (SAS) programming language to the PySpark programming language in a company. With this tool, it is possible to process a large amount of data, taking advantage of the advantages offered by distributed computing and the flexibility of the Python ecosystem. A bibliographical research was carried out and the script developed in SAS for PySpark was transcribed in a technology company in the city of Araraquara - SP. It is concluded that the implementation of the programming logic in PySpark allowed bringing to the corporate environment a tool that has several libraries for data analysis, with agility in processing, reliability in manipulation, which focuses on processing large volumes of data, in addition to facilitating the preparation of studies and the delivery of information.

KEYWORDS: *Analytics. Development. Data. PySpark. Python. SAS.*

RESUMEN
Este trabajo tiene como objetivo migrar el código de un lenguaje de programación del Sistema de Análisis Estadístico (SAS) al lenguaje de programación PySpark en una empresa. Con esta herramienta es posible procesar una gran cantidad de datos aprovechando las ventajas que ofrece la computación distribuida y la flexibilidad del ecosistema Python. Se realizó una investigación bibliográfica y el guión desarrollado en SAS para PySpark fue transcrito en una empresa de tecnología del Municipio de Araraquara - SP. Se concluye que la implementación de la lógica de programación en PySpark nos permitió traer al entorno corporativo una herramienta que cuenta con varias librerías para el análisis de datos, con agilidad en el procesamiento, confiabilidad en la manipulación, que se enfoca en procesar grandes volúmenes de datos, además de facilitar la preparación de estudios y la entrega de información.

PALABRAS CLAVE: *Analítica. Desarrollo. Datos. PySpark. Python. SAS.*

[1] Universidade de Araraquara - UNIARA.
[2] Universidade de Araraquara. Orientadora. Economista e Bacharel em Direito, Docente do Curso de Engenharia de Computação e Sistema de Informação na Universidade de Araraquara- UNIARA. Araraquara-SP.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

1. INTRODUÇÃO

O SAS “Sistema de Análise Estatística”, no mercado desde 1976, é um dos mais reputados sistemas de análises de dados em microcomputadores. Trata-se de um sistema integrado de aplicações para o processamento e análise estatística de dados, consistindo em módulos de Acesso e Recuperação de Dados, Gerenciamento de Arquivos, rotinas de Geração de Gráficos e Geração de Relatórios (UFJF, 2024).

O Spark é um *framework* para processamento de Big Data construído com foco em velocidade, facilidade de uso e análises sofisticadas. Desenvolvido desde 2009 pela Universidade da Califórnia em Berkeley, o Spark tem muitas vantagens como: permite que aplicações em *clusters* Hadoop executem até cem vezes mais rápido em memória e até dez vezes mais rápido em disco se comparado às outras tecnologias como Big Data, o paradigma MapReduce e o Hadoop (Amorim, 2021).

Este trabalho aborda a combinação do Python com o Spark, permitindo que os cientistas de dados analisem dados em larga escala usando um ambiente Python. As principais características do PySpark incluem: dataframe api, integração com bibliotecas python e mllib (Dsacademy, 2024).

O objetivo deste trabalho é transcrever o script em SAS para a linguagem PySpark, a fim de atender um pedido do cliente de uma empresa do setor de tecnologia no Município de Araraquara-SP.

Foi realizada pesquisa bibliográfica e um estudo em uma empresa no Município de Araraquara-SP. Quando a empresa faz a análise de dados em SAS observa-se que o sistema apresentava lentidão com conjuntos de dados muito grandes, especialmente em sistemas locais ou não otimizados. Frente a essa dificuldade, como a empresa poderá melhorar o fluxo de dados sobre investimentos financeiros?

Diante dessa questão, a hipótese dessa pesquisa é que esse processo de migração de dados na empresa será benéfico e eficaz. Pretende-se fazer a migração de dados da linguagem SAS para a linguagem PySpark, uma vez que esta linguagem é mais rápida na manipulação de dados em grande escala e ela tem a vantagem de ser código aberto, denominada Open-Source, pode ser executado em ambientes locais, em *clusters* Hadoop ou em plataformas de nuvem.

Foi proposta a migração do código SAS que a empresa faz com toda manipulação de dados para um novo código, utilizando a linguagem de programação PySpark que é uma API em Python. Para executar o Spark foi oferecido suporte à colaboração entre Apache Spark e Python. Nessa nova transição foi utilizada a sintaxe do PySpark, sendo realizada a mesma lógica que é oferecido com SAS, ou seja, a manipulação dos dados, o merge com as bases de dados recebida e vendida. A próxima etapa foi o novo *script:* a base enviada pelo cliente com informações de entrada como NDOC e DATA realizou o merge com as informações que o cliente solicitou.

A empresa em estudo vende informações sobre a vida financeira, *score* etc. Essas informações foram encaminhadas para o cliente utilizando o novo *script* em PySpark para fazer o merge entre a base de dados enviada pelo cliente e a base de dados da empresa.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

2. REVISÃO BIBLIOGRÁFICA

Esta seção apresenta uma breve revisão bibliográfica sobre os códigos de linguagem de programação como: SAS e PySpark bem como as vantagens e as desvantagens ao utilizar essas ferramentas.

2.1. SAS

O programa trabalha com quatro ações básicas sobre os dados: Acessar, Manipular, Analisar e Apresentar. Pode ser instalado em diversos ambientes operacionais disponíveis no mercado, possuindo portabilidade de programas e arquivos para qualquer um desses ambientes (UFJF, 2024).

Outro aspecto do Sistema SAS é a habilidade de acessar praticamente qualquer formato de dado, em qualquer base. Mesmo bases de dados descontinuadas comercialmente ainda contam com possibilidade de acesso via SAS. O módulo SAS/ACCESS to é o responsável por essa funcionalidade, bastando escolher o adequado. Por exemplo, SAS/ACCESS to Adabas acessa o banco de dados Adabas. SAS/ACCESS to ODBC acessa diversos formatos, todos mapeados através da interface ODBC. Formatos de texto (CSV, TXT etc.) e o próprio formato SAS, são acessados nativamente pelo Base SAS, sem necessidade de nenhum outro módulo (UFJF, 2024).

O SAS é indicado para desenvolvimento de pesquisas com necessidade de análise de grandes bancos de dados, bem como desenvolvimento e aplicação de ferramentas estatísticas avançadas. O projeto prevê a utilização do *software* por toda a comunidade acadêmica, após treinamento dos usuários, oferecido por tutores ligados ao Departamento de Estatística da UFJF (UFJF, 2024).

O SAS é uma suíte de *software* amplamente reconhecida e utilizada para análise estatística e mineração de dados. A popularidade do SAS entre cientistas de dados decorre da sua capacidade notável de lidar com conjuntos massivos de dados, executar análises estatísticas complexas e entregar resultados confiáveis (Ciencia de Dados Brasil, n/d).

As soluções SAS permitem que os usuários realizem gerenciamento de dados, análises estatísticas, modelagem de negócios, construção de relatórios, aprimoramento de qualidade, desenvolvimento de aplicativos, transformação de dados, extração de dados e uma infinidade de outras operações nos dados coletados (JUMP, 2023).

Outra funcionalidade no SAS é a existência do comando LIBNAME onde permite criar atalhos na criação de caminhos apontando para diferentes lugares e pastas na rede, onde possamos trabalhar com múltiplas tabelas, bem versátil (Communities, n/d).

2.2. PySpark

O Apache Spark é uma plataforma de computação distribuída em cluster que foi projetada para ser rápida e de propósito geral. Um cluster Spark pode ter duas, três ou mesmo 4 mil máquinas, aumentando de forma considerável a capacidade computacional e permitindo processar grandes volumes de dados (Dsacademy, 2024).

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

O Spark permite que aplicações em clusters Hadoop executem até 100 vezes mais rápido em memória e até 10 vezes mais rápido em disco, desenvolver rapidamente aplicações em Java, Scala ou Python. Além disso, vem com um conjunto integrado de mais de 80 operadores de alto nível e pode ser usado de forma interativa para consultar dados diretamente do console. Além das operações de Map/Reduce, suporta consultas SQL, streaming de dados, aprendizado de máquina e processamento de grafos (Amorim, 2021).

O PySpark leva todas essas vantagens para o processo de análise de dados usando uma sintaxe própria da Linguagem SQL (através de funções) ou usando SQL padrão ANSI. Ou seja, você cria o processo de análise usando SQL e Python e o Spark se encarrega de executar seu processo em um cluster de forma transparente e super veloz (Dsacademy, 2024).

O PySpark oferece muitas vantagens para empresas que precisam lidar com grandes volumes de dados, como redução de custos, escalabilidade, flexibilidade e eficiência. Ele permite que as empresas processem grandes conjuntos de dados de forma mais rápida e eficiente do que as ferramentas tradicionais, ajudando a acelerar a tomada de decisões baseadas em dados (Awari, 2023).

O PySpark oferece muitas vantagens para empresas que precisam lidar com grandes volumes de dados, como redução de custos, escalabilidade, flexibilidade e eficiência. Ele permite que as empresas processem grandes conjuntos de dados de forma mais rápida e eficiente do que as ferramentas tradicionais, ajudando a acelerar a tomada de decisões baseadas em dados (Awari, 2023).

As vantagens do PySpark em relação a outras ferramentas de análise de dados incluem sua capacidade de processar grandes volumes de dados, compatibilidade com Python, integração com o ecossistema Spark e flexibilidade. Essas vantagens tornam o PySpark uma ferramenta poderosa para empresas que precisam lidar com grandes conjuntos de dados e obter *insights* valiosos para informar suas estratégias de negócios (Awari, 2023)

O PySpark possibilita a utilização de JOIN, que seria une dois ou mais conjuntos de dados, por exemplo duas tabelas que tenham as mesma chave, em caso positivo de ter a mesma chave ele traz essas informações, tornando uma ferramenta com diversas opções de análises, onde podemos ter alguns definição de JOIN, como o Inner join. O inner é o join padrão do Spark e, por conta disso, provavelmente o mais usado. Esta junção cria uma interseção ao unir registros correspondentes e descarta em ambos os lados os registros onde as chaves não correspondem (Robert, 2021).

O PySpark possibilita a criação de DataFrame, é uma estrutura de dados que organiza os dados em uma tabela bidimensional de linhas e colunas, como uma planilha. Os DataFrames são uma das estruturas de dados mais comuns na análise de dados moderna, pois são uma maneira flexível e intuitiva de armazenar e trabalhar com dados (Databricks, n/d).

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

3. DESENVOLVIMENTO

Foi realizado um estudo em uma empresa do setor de tecnologia no município de Araraquara-SP, no qual a análise dos dados foi realizada utilizando a ferramenta SAS. Dessa Ferramenta foram trazidos dados importantes para tomada de decisão na empresa, tais como: saúde financeira, estimativa salarial, dívidas em aberto. A partir desses dados a empresa poderá traçar metas para alavancar crescimento, realizar estudos sobre o comportamento do mercado e proporcionar mais confiança e credibilidade nas decisões para investir.

A empresa em estudo fornece informações sobre opções de investimentos em geral para outras empresas. Neste sentido, se faz necessário que os dados da empresa em estudo sejam obtidos de forma mais rápida e eficiente.

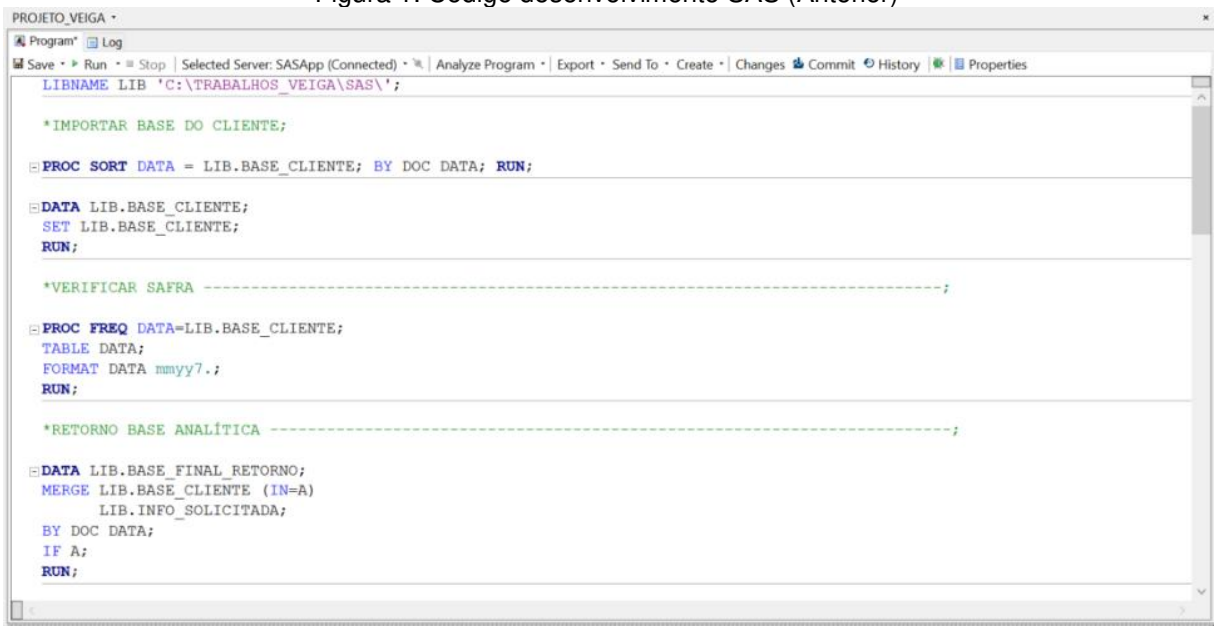
3.1. Utilização do código SAS na empresa

A Ferramenta SAS é utilizada pelos analistas dessa empresa para manipulação de informações comercializada por essa empresa. O funcionamento do sistema no dia a dia é: primeiro o cliente envia uma base enriquecida com informações de usuários: DOC e DATA. Com base nessas informações enviadas é aberto uma requisição para enriquecimento de informações que são vendidas para o cliente. Essas informações são processadas e ao final foi utilizada a ferramenta SAS para fazer o merge dessas informações vendidas com a base de informações que fora enviada pelo cliente. Quando finalizado o merge, o merge é a união das duas tabelas (Tabela enviada pelo cliente BASE_CLIENTE x Tabela de informações solicitadas INFO_SOLICITADA), sendo realizado essa união dos dados comparando o DOC e DATA das duas tabelas e realizando uma nova tabela BASE_FINAL_RETORNO, essa nova tabela com as informações solicitadas são enviadas para o cliente.

A figura 1 apresenta a listagem completa do código SAS. Foi criada a LIBNAME LIB onde foi direcionado o caminho que ficaram todas as bases, sendo elas recebida do cliente BASE_CLIENTE ou bases com informações solicitadas pelo cliente INFO_SOLICITADA. Logo em seguida foi realizada a importação dessa base enviada pelo cliente, direcionando para o LIB que foi criada. Na sequência foi verificada a data enviada pelo cliente. E por fim, realizado o merge entre a BASE_CLIENTE enviada pelo cliente com a base de dados INFO_SOLICITADA, essa base de dados são informações que foram solicitadas pelo cliente para um estudo e análises desses dados. Para realizar o merge, foi preciso ter o mesmo DOC e DATA. No final foi importado a base de dados BASE_FINAL_RETORNO ao cliente com as informações solicitadas.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Figura 1: Código desenvolvimento SAS (Anterior)



Fonte: o autor, 2024.

A figura 2 apresenta o recebimento da base enviada pelo cliente em SAS, com informações de DOC E DATA, denominada BASE_CLIENTE:

Figura 2: Base enviada pelo cliente, desenvolvimento SAS (Anterior)

BASE_CLIENTE

Filter and Sort | Query Builder | Where

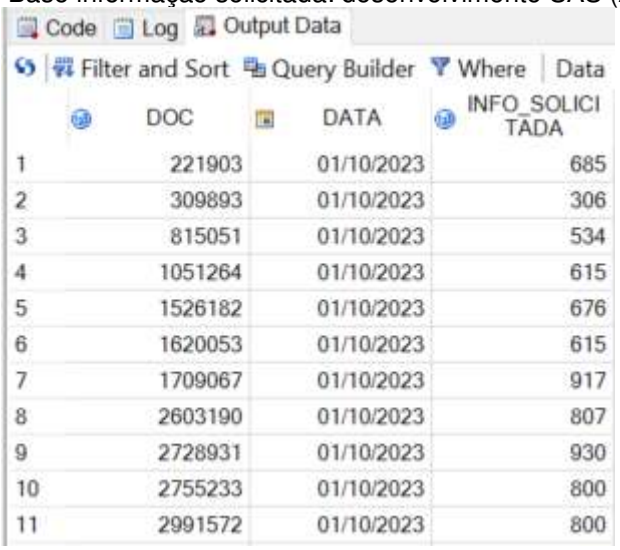
	DATA	DOC
1	01/10/2023	221903
2	01/10/2023	309893
3	01/10/2023	815051
4	01/10/2023	1051264
5	01/10/2023	1526182
6	01/10/2023	1620053
7	01/10/2023	1709067
8	01/10/2023	2603190
9	01/10/2023	2728931
10	01/10/2023	2755233
11	01/10/2023	2991572

Fonte: o autor, 2024.

Na Figura 3 foi exibido o retorno final para o cliente com as informações solicitadas: DOC, DATA, INFO_SOLICITADA.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Figura 3: Base informação solicitada: desenvolvimento SAS (Anterior)



	DOC	DATA	INFO_SOLICITADA
1	221903	01/10/2023	685
2	309893	01/10/2023	306
3	815051	01/10/2023	534
4	1051264	01/10/2023	615
5	1526182	01/10/2023	676
6	1620053	01/10/2023	615
7	1709067	01/10/2023	917
8	2603190	01/10/2023	807
9	2728931	01/10/2023	930
10	2755233	01/10/2023	800
11	2991572	01/10/2023	800

Fonte: o autor, 2024.

3.2. Processo atual do código em PySpark na empresa

Atualmente, a empresa utiliza PySpark para processar informações.

A figura 4 apresenta a listagem completa do código PySpark. O código se inicia com o *import* que foi utilizado durante o processo de manipulação de variáveis com a ferramenta PySpark. A sessão Spark é iniciada para que os comandos sejam obedecidos quando informado para manipularem os dados. Foi criado duas entradas de arquivos: uma, base_cliente_*path* e a outra, info_solicitada_*path*. Na primeira, a informação foi enviada pelo cliente e na segunda, a informação foi vendida pela empresa.

Com as informações carregadas foi criado um *data frame* com nome de base cliente e um segundo data frame foi criado carregando as informações solicitadas com o nome info_solicitada.

Essa leitura no carregamento é feita por *spark read*, onde foi definido o caminho onde se encontra a base e o formato do arquivo que seria .csv. Na sequência foi verificada a safra para o retorno da base analítica final. Foi realizado o merge utilizando comandos Spark como join função da ferramenta e informando qual tipo de merge "inner".

A próxima etapa apresenta informações da base de dados com exibições dos dados quando realizada o merge, para verificação se está tudo correto. Se o merge com as informações foram feitas corretamente, é finalizado o ambiente Spark.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Figura 4: Código desenvolvimento PySpark (Atual)

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, floor
from pyspark.sql.types import IntegerType

# Iniciar a sessão Spark
spark = SparkSession.builder \
    .appName("Transformação de Código SAS para PySpark") \
    .getOrCreate()

# Definir caminhos dos arquivos
base_cliente_path = "C:\\TRABALHOS_VEIGA\\SAS\\base_cliente"
info_solicitada_path = "C:\\TRABALHOS_VEIGA\\SAS\\info_solicitada"

# Carregar dados do cliente
base_cliente = spark.read.format("csv").option("header", "true").load(base_cliente_path)

# Carregar informações solicitadas
info_solicitada = spark.read.format("csv").option("header", "true").load(info_solicitada_path)

# Verificar SAFRA
base_cliente.select("DATA").distinct().show()

# Retorno da base analítica
base_final_retorno = base_cliente.join(info_solicitada, ["DOC", "DATA", "INFO_SOLICITADA"], "inner")

# Exibir schema e algumas linhas da base final de retorno
base_final_retorno.printSchema()
base_final_retorno.show()

# Finalizar sessão Spark
spark.stop()
```

Fonte: o autor, 2024.

A Figura 5 apresenta a base enviada pelo cliente em SAS com informações de DATA e DOC, denominada base_cliente:

Figura 5: Base enviada pelo cliente, desenvolvimento PySpark (Atual)

	DATA	DOC
0	01/10/2023	221903
1	01/10/2023	309893
2	01/10/2023	815051
3	01/10/2023	1051264
4	01/10/2023	1526182
..	...	...
496	01/10/2023	982505661
497	01/10/2023	987728055
498	01/10/2023	987985535
499	01/10/2023	990233423
500	01/10/2023	993127274

Fonte: o autor,2024.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

A Figura 6 apresenta o retorno final para o cliente, com as informações solicitadas: DOC, DATA, INFO_SOLICITADA.

Figura 6: Base informação solicitada: desenvolvimento PySpark (Atual)

	DOC	DATA	INFO_SOLICITADA
0	221903	01/10/2023	685
1	309893	01/10/2023	306
2	815051	01/10/2023	534
3	1051264	01/10/2023	615
4	1526182	01/10/2023	676
..	...	...	...
496	982505661	01/10/2023	823
497	987728055	01/10/2023	731
498	987985535	01/10/2023	353
499	990233423	01/10/2023	708
500	993127274	01/10/2023	712

Fonte: o autor, 2024.

4. RESULTADOS

No processamento dos dados atuais com PySpark foi mantida a estrutura que utiliza a linguagem SAS, trazendo as informações solicitadas (Figuras 7 e 8). O que foi modificado com a transcrição do código não alterou a qualidade e estrutura final da base que enviou/entregou ao cliente.

Figura 7: Base final de entrega para o cliente: desenvolvimento SAS (Anterior)

BASE_FINAL_RETORNO

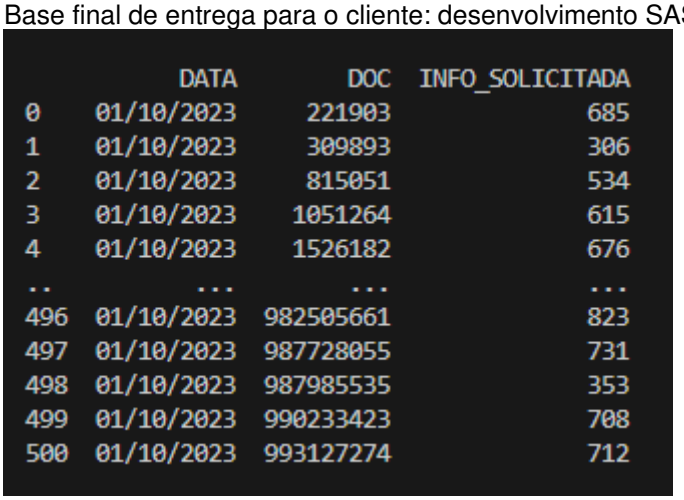
Filter and Sort | Query Builder | Where | Data | D

	DATA	DOC	INFO_SOLICITADA
1	01/10/2023	221903	685
2	01/10/2023	309893	306
3	01/10/2023	815051	534
4	01/10/2023	1051264	615
5	01/10/2023	1526182	676
6	01/10/2023	1620053	615
7	01/10/2023	1709067	917
8	01/10/2023	2603190	807
9	01/10/2023	2728931	930
10	01/10/2023	2755233	800
11	01/10/2023	2991572	800

Fonte: o autor,2024.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Figura 8: Base final de entrega para o cliente: desenvolvimento SAS (Atual)



	DATA	DOC	INFO_SOLICITADA
0	01/10/2023	221903	685
1	01/10/2023	309893	306
2	01/10/2023	815051	534
3	01/10/2023	1051264	615
4	01/10/2023	1526182	676
..	...	...	...
496	01/10/2023	982505661	823
497	01/10/2023	987728055	731
498	01/10/2023	987985535	353
499	01/10/2023	990233423	708
500	01/10/2023	993127274	712

Fonte: o autor, 2024.

5. CONSIDERAÇÕES

A partir do objetivo proposto, que foi a migração de código de uma linguagem para outra na análise dos dados, conclui-se que a transcrição foi realizada respeitando a privacidade dos clientes e mantendo a integridade das operações. O código PySpark desenvolvido possibilitou a implementação com alta qualidade e sofisticação, trazendo para o ambiente corporativo uma ferramenta que possui diversas bibliotecas para análises de dados, com agilidade no processamento, confiabilidade na manipulação, que tem como foco o processamento de volumes grandiosos de dado, além de facilitar a elaboração de estudos e a entrega de informações.

REFERÊNCIAS

AMORIM, L. **Introdução ao Spark com Pyspark**. [*S. l.: s. n.*], 2021. Disponível em: https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/80/346/605?inline=1. Acesso em: 17 abr. 2024

AWARI. **PySpark:** a ferramenta que está revolucionando a análise de dados. [S. l.]: Awari, 2023. Disponível em: https://awari.com.br/pyspark/ Acesso em: 21 maio 2024.

CIENCIA DE DADOS BRASIL. **Estatística Básica com SAS:** Fundamentos para Cientistas de Dados. [S. l.]: Ciência de Dados Brasil, s. d. Disponível em: http://surl.li/bnxayh Acesso em: 23 maio 2024

DATABRICKS. **O que é um DataFrame?**. [S. l.]: Databricks, s. d. Disponível em: http://surl.li/qdqqee Acesso em: 12 nov. 2024.

DSACADEMY. **PySpark – Análise de Dados em Larga Escala e a Interseção com SQL**. [S. l.]: Dsacademy, 2024. Disponível em: http://surl.li/pqjefd Acesso em: 29 abr. 2024.

JUMP. **SAS:** A solução completa para análise de dados. [S. l.]: Jump, 2023. Disponível em:https://jump.tec.br/blog/sas-a-solucao-completa-para-analise-de-dados/. Acesso em: 23 maio 2024.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

PKUSNIARUK. **Dica da semana:** Como ler um arquivo excel utilizando LIBNAME XLSX. [S. l.]: Databricks, s. d. Disponível em: http://surl.li/hpuexc. Acesso em: 12 nov. 2024

ROBERT, Carlos. **Joins em PySpark**. [S. l.]: Data Livre, 2021. Disponível em: https://datalivre.medium.com/joins-em-pyspark-3c1d2773eeb1, Acesso em: 12 nov. 2024

UFJF. **O que é o SAS?.** Juiz de Fora, MG: Departamento de Estatística, s. d. Disponível em: https://www2.ufjf.br/estatistica/eventos-e-projetos/projeto-sas/o-que-e-o-sas/ Acesso em: 11 maio 2024.