

Atividade: Importação e Categorização de Tarefas Abandonadas com PySpark

📅 Início: 03/06/2025 – Terça-feira – tabela teste

📅 Prazo: 09/06/2025 – Segunda-feira – tabela teste

🇧🇷 Demonstração funcional: 10/06/2025 – Terça-feira – tabela produção

🎯 Objetivo da Tarefa (User Story)

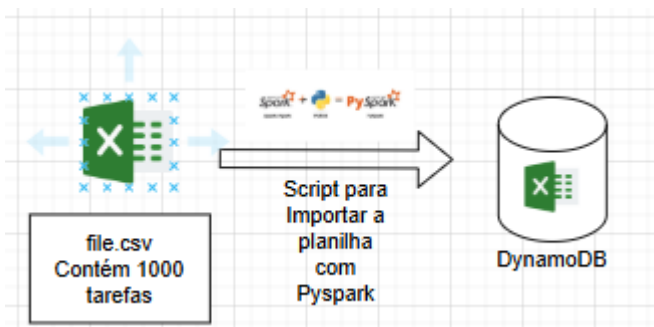
Como nova usuária:

desejo, importar meus dados legados para um sistema novo utilizando PySpark com o objetivo de migrar para um novo sistema, sem perder os dados dele.

🧩 Tarefa 1: Leitura e importação de dados para tabela teste

📋 Etapas:

- Ler a planilha Excel com PySpark (ou Pandas integrado ao Spark).
- Tratar os dados (converter datas, tipos, etc).
- Modelar os itens para estrutura de gravação no DynamoDB.
- Inserir os dados na tabela(teste) DynamoDB via PySpark.
- Testar a persistência dos dados.



📋 Requisitos Funcionais

- Fazer o Sistema continuar funcionando com os dados importados.
- As tarefas precisam ser vinculadas ao novo usuário.

🧪 Critérios de Aceite

1. Importar planilha do Excel com 1000 tarefas com o PySpark.
2. Inserir os dados no banco DynamoDB.

Libs recomendadas:

- Pyspark
- Pandas
- Numpy

Restrições

- Não utilizar IA para desenvolver a solução.
 - Permitido apenas para compreensão de tecnologias como PySpark, Pandas, Numpy etc.

Tarefa 2: Identificação, Categorização e Exportação das Tarefas Abandonadas

Objetivo:

Eu como usuária do sistema

Quero categorizar a cada 6 meses as tarefas abandonadas por mês

para que eu possa ver o volume de tarefas abandonadas e seus tipos por mês

Analisar as tarefas importadas no DynamoDB, identificar as tarefas abandonadas com base nas regras(status da tarefa, data de início e data de conclusão), categorizá-las por mês(lista de compra ou tarefa) e exportar os dados para uma nova planilha Excel.

Etapas:

- Buscar dados no DynamoDB com PySpark.
- Filtrar tarefas abandonadas:
 - Pendentes há mais de 15 dias.
 - Lista de compras com mais de 30 dias.
- Agrupar as tarefas abandonadas por mês de abandono.
- Salvar resultados no DynamoDB (como nova coleção ou item).
- Gerar uma nova planilha com os dados organizados e um relatório

- Exportar o resultado para planilha Excel.

