

Luizalabs

Engenheiro de Dados / Analytics Engineer SR

Bem vindo ao desafio técnico do Luizalabs e muito obrigado por se interessar

em fazer parte do Luizalabs Magalu!

O objetivo deste desafio é avaliar seus conhecimentos técnicos em um contexto

muito similar ao que você encontrará aqui no Magalu.

Para propor as soluções para os desafios abaixo, você terá acesso a alguns arquivos que simulam bases de dados disponíveis em :

[Teste Analytics Engineer 2024.03](https://drive.google.com/drive/u/0/folders/1DBfVSjlpG7QrqxLx1LeTSzSaUwwZQLuE)

Gostaríamos de esclarecer que os dados fornecidos neste link são totalmente fictícios e não correspondem a informações reais. Eles foram criados com o único propósito de possibilitar a resolução de um problema também fictício.

Queremos garantir total transparência em nossa comunicação. Caso você tenha alguma dúvida ou precise de esclarecimentos adicionais, não hesite em entrar em contato.

Contexto Técnico

Temos atualmente no Magalu diferentes serviços que trabalham em conjunto para proporcionar aos nossos clientes, parceiros e vendedores a melhor experiência do Varejo. Em cada um desses microsserviços existem dados que são necessários para seu funcionamento.

Para apoiar os times na tomada de decisões, temos uma Plataforma de Dados que possibilita a ingestão, transformação e visualização de dados.

**Desafio 1**

O Magalu oferta produtos no site, app e loja, mas também oferta serviços, como por exemplo a ***garantia estendida*** *de um produto* e um ***seguro contra roubo e furto***.

Utilize o arquivo *tab\_venda\_servicos.csv* para resolver os desafios 1.1 e 1.2.

Sabe-se que pela coluna ***servico*** dessa tabela em csv é possível identificar o tipo de serviço de acordo com a regra:

* servico = (*5,6,7,8) refere-se a* **'garantia estendida'**
* servico = (19,20) *refere*-se a **‘*seguro* roubo e furto'**

**Desafio 1.1**

Com base nos dados disponibilizados no arquivo mencionado acima, traga como resultado o valor de venda mensal agrupado por mês e por tipo de serviço e valor de venda acumulado por serviço até o mês atual. Salve um arquivo csv de nome ‘desafio1\_1.csv’. com o resultado final.

**Utilize o pyspark na resolução.**

**Desafio 1.2**

João é um colaborador do Magalu que analisa a venda de serviços.

Sempre que um mês é finalizado, João calcula o percentual de crescimento do mês que acabou de finalizar em relação ao mês anterior a este.

O líder do João enviou uma solicitação pra ele:   
*“João preciso enviar para cada filial os 5 meses que tiveram o maior % de crescimento em relação ao mês anterior, você pode gerar uma base que tenha essa informação?”*

Salve um arquivo csv de nome ‘desafio1.2.csv’. com o resultado final.

**Utilize o pyspark na resolução.**

**Output esperado:**

* Um único script python/Jupyter Notebook de nome ‘solucao1’ utilizado para gerar os dois arquivos csv.
  + Não se esqueça que seu código deve ter boas práticas, ser claro e objetivo.
* Resultado do script elaborado
  + O resultado pode ser enviado com a nomenclatura *solucao1.1.csv*’ e ‘solução1.2.csv’

**Desafio 2**

Pense em uma situação que você tenha duas tabelas :

* Tabela\_A com todos os pedidos da empresa contendo as colunas:
  + id\_pedido (int64),
  + data\_pedido (datetime),
  + valor\_pedido (decimal(10,2)),
* Tabela\_B com o evento de todos os pedidos contendo as colunas:
  + id\_pedido (int64),
  + data\_evento (datetime),
  + evento (string),

Exemplos de eventos: ‘*Pagamento aprovado’, ‘Pedido em transporte’, ‘Pedido entregue’, ‘Pedido devolvido’,* etc

Crie uma **consulta SQL** que retorne o id\_pedido, data do pedido, valor do pedido, o último evento de cada pedido e a data deste evento.

**Output esperado:**

* Arquivo solução2.sql contendo a consulta SQL elaborada.

**Desafio 3**

A Fintech Magalu é responsável pela comercialização de serviços financeiros. O colaborador Pedro é analista da fintech e precisa fazer uma análise relacionada a contas, transações bancárias e investimento em renda fixa.

Ele está estudando a possibilidade de oferecer aos clientes um retorno financeiro diário, esse rendimento sempre ocorre ao final de cada dia, após todas as entradas e saídas da conta, se o saldo ao final do dia for negativo não há rendimento, se for positivo gera o rendimento de 0,015% inclusive aos finais de semana.

O arquivo ‘*json\_transacoes\_contas.txt’* disponibilizado no formato JSON, contém dados de transações financeiras por conta referentes ao segundo semestre de 2022. Ele contém o esquema abaixo.

|-- conta: string (nullable = true)

|-- transacoes: array (nullable = true)

| |-- element: struct (containsNull = true)

| | |-- data: string (nullable = true)

| | |-- tipo: string (nullable = true)

| | |-- valor: double (nullable = true)

Sendo:

* **conta**: Identificador da conta (Conta A, Conta B, Conta C , …)
* **transacoes**: Lista de transações de determinada conta
* **data:** Data da transação
* **tipo:** Tipo de transação, podem ser transações de entrada ou de saída
* **valor:** Valor da transação.

Pedro precisa fazer uma análise e necessita de uma tabela com o saldo de cada conta ao final de cada dia, de **01-07-2022** a **31-12-2022**.

Crie um **script python** para ajudar Pedro nessa tarefa, não esqueça de considerar o investimento de acordo com as regras citadas.

**Não utilize linguagem SQL.**

Salve o resultado de um arquivo csv de nome desafio3.csv.

Observe abaixo um exemplo do cálculo como se a taxa fosse de 100%

| Dia | Entrada | Saída | Saldo Diário | Saldo com Investimento |
| --- | --- | --- | --- | --- |
| 1 | 10 | 5 | 5 | 10 |
| 2 | 0 | 12 | -2 | -2 |
| 3 | 20 | 10 | 8 | 16 |

No dia 1 o saldo diário foi positivo, por isso deve um retorno de 100%

No dia 2 o saldo foi negativo, pois tinha 10 na conta (referente ao dia 1) e saiu 12 no dia 2, por isso não teve rendimento.

No dia 3 a conta ficou novamente positiva, e por isso teve rendimento.

**Output esperado:**

* Script python/Jupyter Notebook de nome `solucao4`
  + Não se esqueça que seu código deve ter boas práticas, ser claro e objetivo.
* Resultado do script elaborado
  + O resultado pode ser enviado com a nomenclatura `*solucao4.csv*`
  + O resultado deve ter o seguinte esquema
    - **Colunas**:
      * Conta
      * Data (granularidade por dia)
      * Valor de Entrada (por dia)
      * Valor de Saída (por dia)
      * Saldo Diário (referente às transações do dia)
      * Saldo Diário com Rendimento (Saldo final do dia, com o investimento calculado)

Considerações finais

Responda o email com o resultado do desafio com cópia para [vinicius.kanagusto@luizalabs.com](mailto:vinicius.kanagusto@luizalabs.com) e wagner.soares@luizalabs.com

Deve conter anexo um arquivo .zip com a seguinte estrutura de pastas:

* Desafio\_ae\_{seunomecompleto}.zip
  + Desafio1
    - solução1.py ou solução1.ipynb
    - desafio1\_1.csv
    - desafio1\_2.csv
  + Desafio2
    - desafio2.sql
  + Desafio3
    - desafio3.py ou solução4.ipynb
    - desafio3.csv

Em caso de dúvidas relacionadas ao desafio, não deixe de nos encaminhar um

email com suas perguntas.

Esperamos que se divirta no processo.

Boa sorte!