

Olá, candidato(a)!

Primeiro gostaríamos de agradecer por seu interesse em fazer parte do time de dados da Shopper.com.br.

Queremos conhecer um pouco mais das suas habilidades como Engenheiro de Dados e para isso vamos propor alguns desafios em um cenário fictício.

Seu objetivo será atender os requisitos propostos e apresentá-los da maneira que considerar mais produtiva.

CENÁRIO - Lançamento de um Marketplace Global de Artesanato

Você foi contratado como Engenheiro de Dados para ajudar a lançar um marketplace global de artesanato, onde artistas de todo o mundo podem vender suas criações únicas. Este marketplace visa conectar artesãos com clientes globais, oferecendo uma variedade de produtos, desde joias feitas à mão até móveis artesanais. Sua missão é projetar e implementar uma infraestrutura de dados robusta que suporte todas as operações de negócios, desde o registro de produtos até a análise de vendas e promoções.

Etapa 1 – Modelagem de Dados

Crie o esquema de banco de dados para este marketplace.

Ele deve incluir tabelas para **Clientes** que podem ser tanto compradores quanto vendedores. Um cadastro de **Produtos**, que podem ter atributos variados e ser associados a diferentes **Lojas**, que são gerenciadas pelos vendedores e têm seus próprios perfis, controle de **Estoque** para gerenciar a disponibilidade dos produtos, e um registro de vendas para rastrear todas as transações.

Etapa 2 – Arquitetura de Data Lake

Com o esquema do banco de dados definido, seu próximo passo é desenhar uma arquitetura de um Data Lake em Nuvem que integre o banco de dados do marketplace com fontes de dados externas. Este projeto deve incluir outras duas fontes de dados descritas abaixo:

- **Pesquisas de Preços:** Integração dos dados obtidos por um robô de pesquisa. Os dados são depositados uma vez por dia, em formato CSV, em um diretório SFTP.

Exemplo:

```
product_id,product_name,average_price,minimun_price,maximum_price,survey_date
prod-987654321,Bracelete Artesanal de Prata,150.0,140.0,160.0,2024-04-25
prod-123456789,Vaso de Cerâmica Pintado à Mão,120.0,100.0,140.0,2024-04-26
```

- **Análise de Sentimentos de Redes Sociais:** O marketplace contratou uma API de um fornecedor que monitora redes sociais. Essa API gera um evento toda vez que detecta comentários sobre a marca e seus produtos.

```
{
  "id": "evt-123456789",
  "timestamp": "2024-04-25T14:30:00Z",
  "text": "Adoro as joias deste marketplace! Sempre encontro peças únicas e de ótima
qualidade.",
  "sentiment": {
    "score": 0.95,
    "label": "positivo"
  },
  "product": {
    "id": "prod-987654321",
    "name": "Bracelete Artesanal de Prata"
  },
  "user": {
    "id": "usr-123456789",
    "username": "artlover"
  },
  "source": "Twitter"
},
{
  "id": "evt-987654321",
  "timestamp": "2024-04-26T10:15:00Z",
  "text": "Muito decepcionado com o serviço de entrega deste marketplace, demorou semanas e
o produto chegou danificado.",
  "sentiment": {
    "score": 0.2,
    "label": "negativo"
  },
  "product": {
    "id": "prod-123456789",
    "name": "Vaso de Cerâmica Pintado à Mão"
  },
  "user": {
    "id": "usr-987654321",
    "username": "ceramicfan"
  },
  "source": "Facebook"
}
```

O desenho dessa arquitetura deve representar as fontes de dados, as camadas de ingestão, storage, processamento, datawarehousing, governança, segurança e dataviz.

Considere que o principal provedor de Cloud do marketplace será a GCP, porém sinta-se livre de propor outras ferramentas além das disponíveis no GCP.

Etapas 3 – Transformações de Dados

Utilizando o esquema do banco de dados e a arquitetura de Data Lake desenvolvidos nas etapas anteriores, escreva um código em Python ou SQL (dependendo das suas escolhas de arquitetura) que implemente uma transformação de dados a sua escolha. Ela deve incluir a limpeza de dados, a combinação de dados (como o banco de dados do marketplace, os dados de pesquisa de preços e a análise de sentimentos das redes sociais, por exemplo) e a criação de agregações ou métricas específicas que possam apoiar decisões de negócio.

Atenção: Não é necessário implementar todas as transformações do Lake. Apenas escolha uma que você julgue que vai melhor representar seus conhecimentos

Etapas 4 – Reporting e Data Visualization

Após modelar o banco de dados, integrar diversas fontes no Data Lake e transformar os dados, crie um dashboard para visualizar as métricas chaves de desempenho do marketplace.

Este dashboard deve incluir visualizações de vendas, análise de tendências de produtos mais populares, análise de sentimentos das redes sociais, e competição

Utilize uma ferramenta de BI, como Power BI, Tableau ou Looker, para desenvolver essas visualizações. Além disso, é esperado que ele inclua filtros interativos que permitam aos usuários finais explorarem os dados de diferentes maneiras.

Te desejamos um ótimo trabalho e, caso tenha qualquer dúvida, fique à vontade para entrar em contato conosco.