

Teste Técnico: Pessoa Engenheira de Dados

Involves

Olá candidato(a) ! Tudo bem ? Segue abaixo as questões relacionadas ao teste técnico de Engenharia de Dados da Involves.

Algumas dicas para você se dar bem no teste: Use suas palavras para responder as questões dissertativas; Tire um tempo para se concentrar e fazer o teste da melhor forma possível; Tente não deixar questões sem resposta mesmo que não saiba responder. Iremos avaliar também o esforço do candidato; E por fim, seja criativo, inove ! gostamos bastante dessa skill.

Estaremos na torcida para que você faça um bom teste e seja aprovado(a).

- 1) Descreva com suas palavras os principais conceitos abaixo:**
 - a) O que é um Data Warehouse ?**
 - b) Quais características possuem as tabelas do tipo Fato e Dimensão ?**
 - c) O que é ETL ?**
 - d) Quais são as principais atribuições de um Engenheiro de Dados ?**
 - e) O que é Trade Marketing ?**
- 2) Crie uma query, considerando o SGBD MySQL, para exibir todos os dados de uma tabela de Pontos de Venda (tabela origem PONTO_VENDA_UNIDADE) e restringir apenas os pontos de venda que possuem sell in maior que 20.000 (campo SELLIN) e ainda ordená-los por nome do ponto de venda (campo NOME_PDV).**
- 3) Considerando a tabela de origem da questão anterior, crie uma query que some o valor de sell in de acordo com cada ponto de venda e agrupe os resultados por mês (campo MES) e ano (campo ANO). Ordene os registros por um período cronológico de forma crescente e por nome do ponto de venda.**
- 4) Considerando a tabela de origem da questão 2 e uma segunda tabela VISITAS_PONTO_VENDA, crie uma query que calcule a quantidade de visitas do ponto de venda de nome INVOLVES, sabendo-se que a tabela de visitas possui um campo que identifica se o ponto de venda foi visitado ou não chamado FL_VISITADO (Se 1 = Ponto de venda visitado / Se 0 = Ponto de venda não visitado). O campo chave que liga as duas tabelas é ID_PDV (na tabela PONTO_VENDA_UNIDADE) e FK_PDV(na tabela VISITAS_PONTO_VENDA). A query deve mostrar apenas as informações de nome do ponto de venda e quantidade de visitas realizadas.**
- 5) Considerando a query abaixo, a pessoa engenheira de dados identificou que a performance da query está muito abaixo do esperado. Imaginando que um dos problemas possa estar relacionado aos índices das tabelas do banco de dados, a pessoa resolveu criar os índices nas tabelas. Liste quais possíveis campos devem ser indexados nas tabelas do banco de dados para que a query criada possa performar melhor. Leve em consideração que nenhum campo no banco de dados está indexado.**

```

select
    FT.CICLO,
    FT.ID_DIM_PDV,
    FT.ID_BLOCO_ITEM,
    SUM(FT.QTD_PONTO_EXTRA),
    SUM(FTPI.TOTAL_NOTA_ITEM)
from FT_DOMINANCIA_PONTO_EXTRA_COMPLIANCE FT
inner join TABREF_PAINEL_LOJAS_LP TPLL
    on FT.ID_DIM_PDV = TPLL.ID_DIM_PDV
    and FT.CICLO = TPLL.CICLO
inner join FT_PERFECTSTORE_ITEM FTPI
    on FT.CICLO = FTPI.CICLO
    and FT.ID_DIM_PDV = FTPI.ID_DIM_PDV
    and FT.ID_BLOCO_ITEM = FTPI.ID_BLOCO_ITEM
    and FT.SEMANA_LP = FTPI.SEMANA_LP
where
    FT.CICLO = 202009
    and FT.ID_DIM_PDV = 223459792
group by FT.CICLO,
    FT.ID_DIM_PDV;

```

6) Considere a instrução Python a seguir:

```
x = [ print(i) for i in range(10) if i % 2 == 0 ]
```

Após a execução dessa instrução no Python , a variável “x” conterá qual valor.

7) Faça um script em Python que peça dois números e imprima a soma.

8) Para responder às questões 8, 9 e 10 utilize a ferramenta Pentaho Data Integration (PDI) na versão de sua preferência. A ETL final deve conter um job principal que, por sua vez, deve conter as transformações criadas nas questões 8, 9, 10. Além disso, que tal ganhar um ponto a mais nessas questões ? Para isso, inclua o projeto criado em um repositório do Github (é importante que seja público para termos visibilidade, ok ?). Compartilhe por aqui o link para o repositório.

Segue as questões:

Construa uma transformação que deve usar como datasource o dataset (DATASET_TESTE_DE.csv) que contém informações de coletas de dados nos ponto de vendas. A ETL deve consultar o dataset e inserir, em uma base de dados (modelo dimensional), as informações coletadas, conforme as tabelas abaixo:

- a) Dimensão Calendário (DIM_CALENDARIO): Deve conter data, mês e ano da coleta
- b) Dimensão Ponto de Venda (DIM_PDV): Deve conter o id, nome e perfil do ponto de venda

- c) **Dimensão Linha de Produto (DIM_LINHA_PRODUTO):** Deve conter o id, nome e perfil da linha de produto
- 9) **Construa uma transformação que deve usar como datasource o dataset (DATASET_TESTE_DE.csv) que contém informações de coletas de dados nos ponto de vendas. A transformação deve consultar o dataset e inserir, em uma base de dados (modelo dimensional), as informações coletadas, conforme as tabelas abaixo:**
- a) **Fato Disponibilidade (FT_DISPONIBILIDADE):** Deve conter os ids de ligação das tabelas de dimensões criadas na questão anterior e a quantidade de presenças de cada linha de produto no mês de Setembro/20.
 - b) **Fato Disponibilidade Agregada (FT_DISPONIBILIDADE_AGREGADA):** Deve conter os ids de ligação das tabelas de dimensões (Dimensão Calendário e Ponto de Venda) e a quantidade de presença de linhas de produto agrupadas por ponto de venda no mês de Setembro/20.

Obs: Os dados de “Disponibilidade” estão categorizados na coluna TIPO_COLETA com o valor “Disponibilidade”. A presença é contada sempre que no campo VALOR aparecer o valor “SIM”

- 10) **Construa uma transformação que deve usar como datasource o dataset (DATASET_TESTE_DE.csv) que contém informações de coletas de dados nos ponto de vendas. A transformação deve consultar o dataset e inserir, em uma base de dados (modelo dimensional), as informações coletadas, conforme as tabelas abaixo:**
- a) **Fato Ponto Extra (FT_PONTO_EXTRA):** Deve conter os ids de ligação das tabelas de dimensões criadas na questão anterior e a soma de ponto extras de cada linha de produto no mês de Setembro/20.
 - b) **Fato Ponto Extra Agregada (FT_PONTO_EXTRA_AGREGADA):** Deve conter os ids de ligação das tabelas de dimensões (Dimensão Calendário e Ponto de Venda) e a soma de ponto extras de linhas de produto agrupadas por ponto de venda no mês de Setembro/20.

Obs: Os dados de “Ponto Extra” estão categorizados na coluna TIPO_COLETA com o valor “Ponto Extra”.