TESTE TÉCNICO | ENG DADOS QUALIDADE

QUESTÕES DISSERTATIVAS E DE MÚLTIPLA ESCOLHA

Hernandes Matias Junior

**Questão 1** – Partindo de 2 tabelas para um banco de dados relacional (descriminadas abaixo) não utilizando as funções Min/Max, escreva uma consulta SQL que retorne a lista com a venda de valor mais alto de cada vendedor para o ano de 2016.

**Tabela Vendedor:**

vendedor\_id (int)

vendedor\_nome (string)

**Tabela Venda:**

venda\_id (int)

vendedor\_id (int)

venda\_data (data dd/mm/aaaa)

venda\_valor (double)

**Resposta:**

WITH TRANSACIONAL AS (

SELECT VD.\*,

VDDR. vendedor\_nome,

ROW\_NUMBER() OVER(PARTITION BY VD. vendedor\_id ORDER BY VD.venda\_valor DESC) AS ORDEM

FROM Venda VD

LEFT JOIN Vendedor VDDR

ON VD. vendedor\_id = VDDR. vendedor\_id

)

SELECT T.vendedor\_id,

T. vendedor\_nome,

T. venda\_valor

FROM TRANSACIONAL T

WHERE R.ORDEM = 1

AND YEAR(T. venda\_data) = 2016

**Questão 2** – Prevendo sobreviventes do Titanic com Machine Learning Supervisionado

Desenvolvido no notebook

**Questão 3** – Análise exploratória dos sobreviventes do Titanic

Modelagem desenvolvida no notebook e resultados no material em PPT.

**Questão 4** –Descreva os principais pontos que definem os conceitos ELT assim como sua aplicação

**Resposta:**

**ETL** é a sigla para “extração, transformação e carregamento”, etapas realizadas em sequência a fim de reunir dados de fontes diferentes; transformá-los em um formato em que possa ser analisado; e finalmente armazená-los em outro banco de dados. O processo de ETL é comumente utilizado na construção de um data warehouse, onde dados de diferentes fontes (SAP, Salesforce, Sydle) precisam ser padronizados para poderem ser analisados posteriormente. Já a ELT é uma inversão das duas últimas etapas, onde o carregamento no banco-destino acontece antes da transformação. Essa inversão faz com que a etapa de transformação seja feita dentro do próprio data warehouse, aproveitando o desempenho do próprio banco e reduzindo o tempo de carregamento dos dados.

**Questão 6** – O que faz o código SQL abaixo?

select data\_transacao,

count(case when cliente\_ativo = 'S' then 1 else 0 end) as qtde\_de\_clientes,

sum(case when cliente\_ativo = 'S' then valor\_transacao else 0 end) as valores\_transacionados

from transaction\_table

group by data\_transacao

;

a) Consolida a quantidade de clientes por valor da transação

b) Consolida a quantidade de clientes por valor e data da transação

c) Consolida a quantidade de clientes ativos e o valor transacionado por data de transação

d) Consolida a quantidade de clientes ativos e não ativos e o valor transacionado por venda

e) Consolida a quantidade de clientes ativos por data de transação

**Questão 7** – Quando uma empresa utiliza Data Warehouse (DW) é necessário carregá-lo para permitir a análise comercial. Para isso, os dados de um ou mais sistemas devem ser extraídos e copiados para o DW em um processo conhecido como:

a) ERP

b) BI

c) CRM

d) ETL

e) Data Mart

**Questão 8** – Para projeto de implementação de DW um dos modelos mais utilizados é o modelo dimensional ou multidimensional. Assinale qual a alternativa abaixo melhor corresponda a um modelo dimensional (composto por uma tabela fato e várias tabelas dimensão):

a) As tabelas dimensão devem conter apenas atributos do tipo literal

b) A tabela fato tem uma cardinalidade de mapeamento de um para um com cada tabela dimensão

c) A tabela fato deve conter atributos numéricos, visando proporcionar dados para uma análise de atividades da empresa

d) Há um número teórico mínimo de 3 e máximo de 15 tabelas dimensão

e) As tabelas dimensão comportam um número máximo teórico de atributos

**Questão 9** – Quais dos itens abaixo não são considerados banco de dados relacional?

a) SQL Server

b) Oracle

c) SAS

d) MySQL

e) Hadoop

f) Microsoft Access

g) Cassandra

**Questão 10** – Em relação ao DW e conceitos associados, é correto afirmar:

a) Ferramentas OLAP podem operar sobre bases de dados armazenadas em SGBDs orientados a objetos. Esta forma de armazenamento é conhecida como OOLAP ou Object OLAP. Também podem operar em bases relacionais, conhecidas como ROLAP ou Relational OLAP, que não são multidimensionais

b) Os maxicubos apresentam a característica de possuir armazenamento e indexação em estruturas de dados que otimizam atualizações ao invés de consultas. Quando o modelo multidimensional é processado, nova base é gerada, contendo somente as agregações, em formato próprio, utilizando-se de estruturas apropriadas para pesquisas

c) Uma das formas de apresentação de um banco de dados multidimensional é através do Modelo SnowStar ou Estrela de Neve. É um modelo simples e eficiente, caracterizado por possuir uma única tabela de fatos e chaves simples nas tabelas de dimensões. Cada dimensão é representada por uma única tabela

d) Com relação à granularidade, as bases de dados operacionais trabalham com o maior nível de detalhe possível, ou seja, maior granularidade. Já no DW pode haver diversos graus de agregação e resumo dos dados. A correta determinação da granularidade exerce papel fundamental no planejamento de capacidade e desempenho do DW

e) Ao contrário do que ocorre com as bases operacionais, o DW, por conter dados históricos, demanda alta taxa de atualização. Desse modo, deve ser atualizado a cada 4, ou no máximo, 24 horas. Além disso, sofre muitas modificações, embora de forma controlada