

OLAP

É uma técnica que permite análise interativa e multidimensional de grandes conjuntos de dados. Ele facilita consultas complexas e permite aos usuários explorar os dados de diferentes formas

Uma empresa usa o OLAP para analisar vendas por região, produto e tempo

Data warehouse

É um repositório centralizado de dados que são extraídos, transformados e carregados de várias fontes e é projetado para armazenar grandes volumes de dados históricos e atuais.

Uma empresa armazena dados de faturamento, chamadas e clientes para análise de tendências

KDD

Refere-se ao processo de descoberta de conhecimento em grandes conjuntos de dados, envolvendo etapas como seleção, pré-processamento, mineração de dados e interpretação dos resultados.

Uma empresa pode usar para entender o comportamento de clientes, para descobrir padrões de compra, segmentar clientes personalizar ofertas entre outros

Data mart

É um subconjunto de data warehouse, focado em área específica ou departamento da empresa. Ele contém dados selecionados e pré-processados para atender as necessidades de um grupo específico de usuários.

Um datamart de marketing contém dados de comportamento de clientes para estratégias personalizadas.

Data mining

É o processo de descoberta de padrões, tendências ou informações úteis a partir de grandes conjuntos de dados usando técnicas estatísticas, machine learning e inteligência artificial

Um banco usa data mining para identificar atividades suspeitas em transações e detectar fraudes

2) sobre o esquema de estrela, utilizado na modelagem de data warehouse responda

a) qual a diferença entre tabelas de fatos e dimensões, explique em um esquema

tabelas de fatos: armazenam dados quantitativos em um esquema de datawarehouse, contendo chaves estrangeiras para tabelas de dimensões e medidas numéricas

tabelas de dimensões: armazenam atributos descritivos ou categorias utilizadas para analisar dados nas tabelas de fatos, sendo referência para chaves estrangeiras

exemplo de esquema:

-Tabela de fatos (vendas): chaves estrangeiras para dimensão de tempo de produto, além do valor da venda

-Tabela de dimensão (dimensão de tempo e produto): armazenam atributos descritivos associados ao tempo e produtos.

b) por que normalmente esse tipo de esquema não é normalizado?

-Normalização completa é evitada para otimizar desempenho em consultas analíticas

-A desnormalização simplifica consultas, facilita a compreensão e manutenção, permite armazenar dados eficientes

-A propriedade é a eficiência em consultas analíticas, justificando a opção pela desnormalização em esquemas de data warehouse

3) a) Mostre os nomes das aeronaves tais que todos os pilotos certificados para operá-las ganhem mais do que R\$ 50.000,00.

```
SELECT DISTINCT A.nomeAeronave
FROM AERONAVE A
WHERE NOT EXISTS (
    SELECT F.numFunc
    FROM FUNCIONARIOS F, CERTIFICADO C
    WHERE F.numFunc = C.numFunc AND C.numAeronave = A.numAeronave
    AND F.salario <= 50000
```

b) Mostre o nome da aeronave e o salário médio de todos os pilotos certificados a voar em aeronaves com distância Max maior do que 1000 quilômetros.

```
SELECT A.nomeAeronave, AVG(F.salario) AS salario_medio
FROM AERONAVE A, CERTIFICADO C, FUNCIONARIOS F
WHERE
    A.numAeronave = C.numAeronave
    AND C.numFunc = F.numFunc
    AND A.distanciaMax > 1000
GROUP BY
    A.nomeAeronave;
```

4) Considerando o benchmark TPC-H escreva as seguintes consultas em SQL.

a) Relatório que mostre o nome e o telefone dos clientes (Customer), além do somatório das quantidades de itens (Lineitem) comprados por eles, ordenado de maneira decrescente por este somatório.

```
SELECT
    c.nome AS nome_cliente,
    c.telefone AS telefone_cliente,
    SUM(i.quantidade) AS total_itens_comprados
FROM
    Clientes c
JOIN
    Pedidos p ON c.id_cliente = p.id_cliente
JOIN
    ItensPedido i ON p.id_pedido = i.id_pedido
GROUP BY
    c.id_cliente, c.nome, c.telefone
ORDER BY
    total_itens_comprados DESC;
```

b) Relatório que apresente a quantidade total de itens (Lineitem) de todas as encomendas agrupadas, bem como a sua quantidade média de itens e a média do preço cobrado/estendido.

```
SELECT
    SUM(quantidade) AS total_itens,
    AVG(quantidade) AS media_quantidade,
    AVG(preco_cobrado) AS media_preco_cobrado
FROM
    ItensPedido;
```

c) Relatório dos meios de transporte (shipmode) agrupados, mostrando a quantidade total, somatório e a média das taxas (tax) pagas em ordem alfabética pelo meio de transporte.

```
SELECT
    shipmode,
    COUNT(*) AS quantidade_total,
    SUM(tax) AS somatorio_taxas,
    AVG(tax) AS media_taxa
FROM
    TabelaEnvios
GROUP BY
    shipmode
ORDER BY
    shipmode;
```

d) Relatório que apresente o saldo médio dos clientes (Customer) de determinada nação (Nation) e região (Region). A consulta deverá agrupar as regiões e nações e apresentar resultado ordenado por região.

```
SELECT
    shipmode,
    COUNT(*) AS quantidade_total,
    SUM(tax) AS somatorio_taxas,
    AVG(tax) AS media_taxa
FROM
    TabelaEnvios
GROUP BY
    shipmode
ORDER BY
    shipmode;
```