

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO  
DISCIPLINA: ARMAZÉM DE DADOS  
TESTE DE AVALIAÇÃO (Valor: 100 pontos)  
Nome:

Observações Gerais:

1. Duração da prova: 3 horas
2. A interpretação das questões faz parte da prova
3. A prova é individual e sem consulta
4. Boa sorte!

1ª. Questão (15 pontos)

Explique as principais motivações para a construção de um Data Warehouse e quais as principais dificuldades de se construir um. Explicitamente indique quais as dificuldades de se construir um Sistema de Suporte a Decisões a partir de sistemas operacionais (ou em operação) do negócio e os benefícios que o DW pode trazer.

2ª. Questão (15 pontos)

Utilizando com um exemplo **não discutido em sala de aula** (i.e., que não está nas transparências) explique porque o modelo dimensional pode ser considerado altamente **extensível**, tanto do ponto de vista de dimensionalidade quanto de fatos analisados. Explique como essa extensibilidade pode ser realizada numa implementação relacional

3ª. Questão (15 pontos)

Explique o dilema do armazenamneto (ou não) dos agregados de em DW. Trace um paralelo dessa questão com as estratégias de implementação de visões em bancos relacionais (materialização da visão *versus* modificação da consulta)

4ª. Questão (15 pontos)

Explique porque as estratégias para acomodar as mudanças que ocorrem mais lentamente em um data warehouse (e.g., endereço de cliente, descrição de produto) não seriam adequadas para mudanças rápidas e freqüentes em dimensões “monstruosas” (grande volume de dados) com mudanças freqüentes.

5ª. Questão (10 pontos)

Desenvolva um modelo dimensional para um Data Mart cujo objetivo é analisar e comparar a produção científica de vários professores e departamentos de Ciência da Computação. O fato em si é a publicação de um artigo em algum evento científico ou em algum periódico num ano específico. Um artigo científico pode ter vários autores (numero máximo não pode ser previamente determinado) e é preciso indicar todos os autores para cada artigo.

6ª. Questão (15 pontos)

Considere uma dimensão cliente contendo informação sobre centenas de milhares de clientes onde dentre as várias informações do cliente existem informações demográficas tais como nível de educação, número de filhos, limite de crédito, etc, que podem mudar frequentemente. Considere que existam 10 campos relacionados ao cliente que não mudam e 5 campos com informação demográfica que pode mudar com frequência. Considere que a cada dia são alteradas em média 100 entradas de clientes e que cada alteração implica numa nova linha na dimensão cliente para representar a nova situação do cliente. Estime quanto espaço em disco será gasto ao final de 1 mês (considere que cada campo consome em média 8 bytes). Explique como uma estratégia baseada em mini-dimensões para os atributos demográficos pode ajudar a amenizar o problema de espaço e custo de atualização e estime quanto de espaço poderia ser economizado. Considere que cada um dos cinco atributos demográficos pode assumir 4 valores.

7ª. Questão (15 pontos)

Explique a importância das dimensões e fatos conformes para a construção de um DW consolidado a partir de data marts individuais. Dê um exemplo concreto **não discutido em sala de aula (i.e., que não está nas transparências)** dessa importância e do tipo de análise que pode ser feito a partir de fatos e dimensões conformes considerados **dois** data marts conformados (ou compatíveis). O exemplo precisa considerar análises que não poderiam individualmente ser realizadas com os dois data marts primários, isto é, precisa mostrar o valor do DW corporativo (construído a partir dos data marts individuais).