

Análise de Dados com Base em Processamento Massivo em Paralelo

Lista de Exercícios: Introdução

Profa. Dra. Cristina Dutra de Aguiar

Observação:

Esta lista contém exercícios classificados como essenciais e complementares. A indicação da classificação de cada exercício é feita junto de sua definição. Recomenda-se fortemente que a lista de exercícios seja respondida antes de se consultar as respostas dos exercícios.

1. (Essencial) Qual a diferença entre OLTP e OLAP?
2. (Essencial) Liste os principais aspectos pelos quais o ambiente de *data warehousing* se difere do conceito de *data warehouse*.
3. (Essencial) Na aula, foram contextualizados dois ambientes: o ambiente operacional e o ambiente informacional. Esses ambientes são distintos entre si e cada um deles tem características importantes que os definem. Descreva cada um desses ambientes, destacando as suas principais características.
4. (Essencial) O reitor de uma universidade precisa tomar uma decisão acerca da distribuição de recursos financeiros entre os institutos da universidade. Em reunião com os pró-reitores da instituição, foi definido que aqueles institutos com maior quantidade de publicações científicas em revistas de alto fator de impacto devem ser priorizados. Sendo assim, o reitor decidiu utilizar a ferramenta OLAP da universidade para emitir um relatório contendo a quantidade de publicações por instituto, por revista e por trimestre. O relatório emitido pelo reitor pode ser classificado em qual categoria: dado, informação ou conhecimento? Justifique sua resposta detalhando o porquê do relatório não ter sido classificado nas outras duas categorias.
5. (Essencial) Considere uma empresa de supermercados que possui várias filiais. Cada filial possui um sistema diferente para contabilizar os produtos vendidos e as promoções realizadas. Um executivo dessa empresa deseja fazer uma análise para descobrir filiais que precisam ser fechadas ou remodeladas. Por que não é ideal que esse executivo realize essa análise sobre os sistemas existentes da empresa?

6. (Complementar) Considere a seguinte situação:

Uma empresa brasileira especializada em análise de dados busca estudar e entender as maiores consequências que a pandemia gerada pelo novo coronavírus ocasionou nas cidades brasileiras. O principal objetivo é determinar quais diferentes medidas podem ser tomadas, levando em consideração as diversas características socioeconômicas que cada cidade ou região do país pode ter. Para dar início ao estudo, a empresa catalogou conjuntos de dados que contêm índices socioeconômicos e dados referentes aos índices de contaminação, quantidade de testes, recuperação e óbitos para cada cidade.

Com base nessa descrição, cite quais dados podem ser extraídos, quais informações podem ser obtidas e quais conhecimentos podem ser construídos.

7. (Complementar) Pesquise uma situação real na qual seria interessante aplicar consultas analíticas. Descreva o problema e elabore três perguntas para exemplificar possíveis análises. A partir dessas perguntas, descreva três exemplos de conhecimento que podem ser gerados por meio dessas análises.

