

Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Departamento de Ciências de Computação
Processamento Analítico de Dados (em Larga Escala)
Lista de Exercícios da Matéria Introdutória

1. Qual a diferença entre OLTP e OLAP?
2. Qual a diferença entre *data warehouse* e *data warehousing*? Qual a relação existente entre esses conceitos?
3. Considere as seguintes expressões:
 - a) inserção/remoção/atualização
 - b) OLAP
 - c) grande número de usuários concorrentes
 - d) interação dinâmica
 - e) OLTP
 - f) interação estática/predefinida
 - g) poucos usuários concorrentes
 - h) muitos registros acessados por vez
 - i) leitura (consulta)
 - j) poucos registros acessados por vez

Preencha a tabela adequadamente utilizando essas expressões.

Ambiente Operacional	Ambiente Informacional

4. O que é o processo de ETL?
5. Por que o *data warehouse* é considerado o “coração” do *data warehousing*?
6. Quais as características dos dados do *data warehouse*? Cite e explique cada uma dessas características.

Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Departamento de Ciências de Computação
Processamento Analítico de Dados (em Larga Escala)
Respostas dos Exercícios da Matéria Introdutória

1. OLTP (*on-line transaction processing*) diz respeito ao ambiente operacional, voltado ao processamento de transações. Isso significa que no ambiente OLTP existem muitas operações de inserção, remoção e atualização e que o objetivo de desempenho é realizar o processamento eficiente dessas operações.

OLAP (*on-line analytical processing*) diz respeito ao ambiente informacional, voltado ao processamento de consultas analíticas. Isso significa que no ambiente OLAP existem muitas consultas e que o objetivo de desempenho é realizar o processamento eficiente dessas consultas.

2. *Data warehouse* representa o banco de dados, ou seja, é o local onde os dados são armazenados. *Data warehousing*, por sua vez, representa um ambiente, o qual é composto por *data warehouse*, *software*, *hardware* e *peopleware*.

Data warehouse é um dos componentes de maior importância do *data warehousing*, consistindo no local onde os dados resultantes do processo de ETL (*extract, transform, load*) e modelados multidimensionalmente são armazenados.

3. Tabela com as respostas:

Ambiente Operacional	Ambiente Informacional
inserção/remoção/atualização	leitura (consulta)
OLTP	OLAP
interação estática/predefinida	interação dinâmica
poucos registros acessados por vez	muitos registros acessados por vez
grande número de usuários concorrentes	poucos usuários concorrentes

4. No processo de ETL (*extract, transform, load*), dados de interesse das fontes de dados são extraídos, transformados e carregados no *data warehouse* ou nos *data marts*. A letra E se refere à extração de dados das fontes de dados. A letra T se refere às transformações sofridas pelos dados, e inclui várias operações: tradução entre

diferentes formatos e modelos, integração de dados heterogêneos, limpeza dos dados.

A letra L se refere à carga dos dados no *data warehouse* ou nos *data marts*.

5. O *data warehousing* pode ser considerado o “coração” do *data warehousing* porque ele armazena dados que já passaram pelo processo de ETL. Esses dados são caracterizados por serem estruturados e por estarem organizados multidimensionalmente, de acordo com as diferentes perspectivas de análise dos usuários de sistemas de suporte à decisão. Utilizando o *data warehouse* como base, as consultas analíticas (consultas OLAP) podem ser respondidas eficientemente.
6. Os dados do *data warehouse* são:
 - a) Orientados a assunto: voltados aos assuntos de maior interesse da corporação.
 - b) Históricos: relativos a um longo período de tempo.
 - c) Integrados: podem ser obtidos de diferentes fontes de dados, as quais possuem diferentes formatos e modelos. Porém, passam por processos de ETL que realizam a tradução dos dados para um modelo comum e a limpeza dos dados, além da integração dos dados para remover conflitos entre os dados.
 - d) Não-voláteis: permanecem estáveis por longos períodos de tempo. Novas inserções são caracterizadas por serem do tipo *append-only*. Adicionalmente, o ambiente é voltado ao processamento de consultas (*read-only*).