

Análise de Dados com Base em Processamento Massivo em Paralelo Arquitetura de Data Warehou- sing: Respostas dos Exercícios

Profa. Dra. Cristina Dutra de Aguiar Ciferri

André Perez

Guilherme Muzzi da Rocha

Jadson José Monteiro Oliveira

João Pedro de Carvalho Castro

Leonardo Mauro Pereira Moraes

Piero Lima Capelo

Observação:

Recomenda-se fortemente que a lista de exercícios seja respondida antes de se consultar as respostas dos exercícios.

1. 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*. Ou seja, no processo de ETL, os dados são transformados antes de ocorrer a carga no *data warehouse* ou nos *data marts*.

No processo de ELT (*extract, load, transform*), dados de interesse das fontes de dados são extraídos e armazenados no *data lake* sem passar por transformações. Na sequência, os dados são extraídos do *data lake*, transformados e carregados no *data warehouse* ou nos *data marts*. Ou seja, no processo de ELT, os dados são extraídos e carregados no *data lake* sem sofrer transformações. Essas transformações somente acontecem depois, quando os dados forem armazenados no *data warehouse* ou nos *data marts*.

2. 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.

3. Conjuntos de dados e seus respectivos componentes:

- Repositório de Metadados: Nenhum conjunto de dados.
- *Data Lake*: Conjuntos (b) e (e).
- *Data Warehouse*: Conjunto (d).
- Bancos de Dados Operacionais: Conjunto (f).
- *Data Staging Area*: Conjunto (a).
- *Data Marts*: Conjunto (c).

4. Tabela preenchida:

Data Warehouse	Data Lake
dados consolidados, organizados e estruturados	dados estruturados, semiestruturados e não estruturados
latência alta	latência baixa
menor custo de análise	maior custo de análise
dados pré-processados antes de serem carregados	dados extraídos e carregados, sem sofrer transformações
esquema estruturado (formato bem definido)	esquema em formato nativo (diferentes formatos)
maior custo de geração dos dados	menor custo de geração dos dados
consultas OLAP	tipos de consulta variados
ETL	ELT

5. Volume, Velocidade, Variedade e Visualização.

6. A Proposta 1 e a Proposta 3 são ferramentas de construção de *dashboards* interativos de *software* livre e gratuitas, sendo compatíveis com a tecnologia do DW e servidor OLAP AWS Redshift. Portanto, pode-se escolher uma dessas duas opções.
7. Opção (a). O motor de consulta deve se conectar ao *data lake* para facilitar a execução de consultas executadas contra os dados armazenados nessa área de armazenamento. O *data mart* deve se conectar ao DW, desde que seus dados são um subconjunto dos dados armazenados no DW.

