

Unidade Curricular: Sistemas de Informação II

Ficha T. Prática n.º 2

Objectivo:

Reflectir sobre a definição de DW.

Texto para reflexão

Segundo a Wikipedia,

*"A **data warehouse (DW)** is a database used for reporting. The data is uploaded from the operational systems for reporting. The data may pass through an [operational data store](#) for additional operations before it is used in the DW for reporting.*

A data warehouse maintains its functions in three layers: staging, integration, and access. Staging is used to store raw data for use by developers (analysis and support). The integration layer is used to integrate data and to have a level of abstraction from users. The access layer is for getting data out for users.

This definition of the data warehouse focuses on data storage. The main source of the data is cleaned, transformed, catalogued and made available for use by managers and other business professionals for [data mining](#), [online analytical processing](#), market research and decision support (Marakas & OBrien 2009). However, the means to retrieve and analyze data, [to extract, transform and load](#) data, and to manage the [data dictionary](#) are also considered essential components of a data warehousing system. Many references to data warehousing use this broader context. Thus, an expanded definition for data warehousing includes [business intelligence tools](#), tools to extract, transform and load data into the repository, and tools to manage and retrieve [metadata](#)."

Já Bill Inmon, define o DW como:

"A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process".

Por seu lado, Ralph Kimball refere uma definição muito mais simples no seu livro "The Data Warehouse Toolkit", 1.ª edição:

A data warehouse is "a copy of transaction data specifically structured for query and analysis".

Tarefas propostas:

1. Procure enquadrar no texto inicial, a primeira definição de DW (segundo Bill Inmon), justificando cada enquadramento efectuado.
2. O mesmo para a segunda definição.
3. No final do 1.º texto, surge, no âmbito do Data Warehouse, a questão da catalogação, a gestão do dicionário de dados e metadados, parecendo ser assunto de importância central na arquitectura. Discuta a sua aparente relevância e, especialmente, a sua utilização dual.
4. A definição de Bill Inmon parece encerrar, em si, uma contradição. Como podemos ter dados não voláteis, mas variantes no tempo? Discuta a aparente contradição.
5. Apresente e explique a arquitetura genérica de uma *Data Warehouse*, identificando os principais elementos que a compõem.
6. Quais os perigos de utilizar um DW não integrado.
7. Examine os benefícios de introduzir um Data Warehouse numa organização como suporte ao processo de tomada de decisões, bem como os potenciais obstáculos ou desafios que podem surgir durante a criação e manutenção de uma estrutura de Data Warehouse.
8. De que forma a arquitetura de um Data Warehouse pode ser personalizada para satisfazer as necessidades específicas de uma organização, tendo em conta as dimensões como o volume e a diversidade dos dados?
9. Quais são os principais desafios na integração de fontes de dados heterogéneas em um ambiente de Data Warehouse e como podem ser superados.
10. Quais são os métodos e tecnologias comuns utilizados para realizar consultas e análises em um ambiente de Data Warehouse, e como esses processos contribuem para a tomada de decisões eficaz em uma organização?