

Ficha T. Prática nº2

Exercício 1:

De acordo com a definição do autor Bill Inmon e, enquadrando com o texto inicial, a Data Warehouse (DW) é uma base de dados usada para relatórios. A definição diz de Bill Inmon diz que uma Data Warehouse (DW) é um conjunto de dados orientados por assunto, integrados, históricos e não voláteis. Isto significa que é projetado para ser um repositório de dados que atende às necessidades de relatórios e análises de uma organização. As características definidas na definição garantem que a Data Warehouse (DW) é uma fonte de informação de confiança para apoiar as tomadas de decisão na área da gestão. O ponto de contacto das duas perspetivas do DW acaba por ser a existência do repositório de dados, e a perspetiva arquitetural e a sua utilização para o negócio.

Exercício 2:

De acordo com a definição do autor Ralph Kimball e, enquadrando com o texto inicial, a Data Warehouse (DW) é uma cópia dos dados de transações especificamente estruturada para consulta e análise. Assim, o Data Warehouse (DW) serve como uma fonte de dados para apoiar as necessidades e análises de negócios. A definição do autor é bastante mais simples. A cópia de dados oriundos dos sistemas transacionais surge logo no início do texto “uploaded from the operational systems for reporting”. Mais à frente, surge a falar na necessidade de existência de meios. No final, refere-se à existência de “tools to extract, transform and load data into the repository, and tools to manage and retrieve metadata.”, onde está subjacente a extração de dados de fontes. Em relação à natureza especificamente estruturada, não é focada especificamente no texto, mas é implícita ao falar-se em “transform data” em duas fases do texto. Por último, a questão de utilização de dados, está clara em “and made available for use by managers and other business professionals for data mining, online analytical processing, market research and decision support”.

Exercício 3:

A relevância destes temas pode ser explicada devido ao papel fundamental que desempenham nas operações de uma Data Warehouse (DW). A catalogação envolve a organização e documentação dos dados da Data Warehouse (DW), que inclui a criação de metadados que descrevem os dados. A gestão do dicionário de dados e metadados assegura que os dados estejam corretamente documentados, para que haja uma garantia da qualidade dos mesmos, evitando assim eventuais erros. A catalogação e os metadados ajudam também os utilizadores a entender a Data Warehouse (DW), pois fornecem informações sobre os dados. Uma boa gestão do dicionário de dados permite uma boa recuperação de eventuais perdas de dados, permitindo também aos utilizadores uma melhor seleção e procura

de dados através de descrições e metadados. Além disto, facilita o desenvolvimento de processos ETL através dessas informações.

Correção (Exercício 3):

A questão dos metadados e a sua gestão é de primordial importância. Realmente, um Data Warehouse, com o previsível crescimento contínuo, acaba por se tornar algo complexo, e, especialmente, tendo de acomodar sucessivas alterações na sua estrutura, operações de extração e transformação e no próprio significado de cada dado, em termos do negócio. Assim, a questão da catalogação de cada artefacto de dados, com inclusão de toda a Informação possível sobre ele, é de primordial importância, surgindo a utilização de ferramentas para suportarem essa tarefa. De salientar que os metadados, têm uma utilização dual, aumentando assim ainda, a sua Importância: Não só encerram uma perspectiva técnica, permitindo a automatização do processo de ETL e a sua manutenção (algo patente no texto acima em "Staging is used to store row data for use by developers (analysis and support)." e em "tools to extract, transform and load data into the repository, and tools to manage and retrieve metadata.", falando-se nos metadados, ferramentas para a sua gestão e busca e sua utilização em termos da área de estágio, por parte dos developers. Mas a sua utilização em termos das ferramentas de acesso é relevante: estas, identificam os dados disponíveis e o seu significado para o negócio e podem assim apresentá-los ao utilizador, escondendo os detalhes técnicos e construtivos. Esta realidade, é mostrada no texto em "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.". Realmente refere-se que a área de integração permite um nível de abstração para os utilizadores (pondo à disposição da camada de acesso os dados, no modo nativo como os utilizadores vêem o negócio). Ora isto é conseguido através dos metadados e do conhecimento que permitem do significado de cada dado para o negócio.

Exercício 4:

Esta contradição entre “dados não voláteis” e “variantes no tempo” pode dever-se ao verdadeiro significado destas duas características. Esta contradição é resolvida reconhecendo que os dados da Data Warehouse (DW) são não voláteis no que diz respeito a alterações posteriores, mas são variantes no tempo para suportar análises históricas e capacidade de rastreamento da evolução de dados.

Correção (Exercício 4):

Esta parte da definição tem subjacente a característica de possíveis "snapshots temporais" do DW. Realmente, num DW os dados são caracteristicamente não voláteis, ou seja, uma vez carregados no repositório, não são posteriormente alterados, a não ser em casos de deteção de erros ou na atualização de agregações materializadas (umas e outras de reduzido âmbito). Uma outra situação de alteração surgiria, com a remoção de dados relativos a períodos que, porventura, tenham perdido

valor, em termos analíticos. Mas, os dados do DW encerram em si uma perspectiva temporal. Podemos vê-los como uma sequência de fotografias temporais (tipo filme). Cada fotografia mostra a realidade da empresa em cada momento e, assim, permite-se a análise dos vários parâmetros de funcionamento do negócio ao longo do tempo. Cada fotografia é imutável, mas a sua sucessão mostra a evolução da empresa no tempo.

Exercício 5:

Os principais componentes da arquitetura de um Data Warehouse (DW) são: Fontes de Dados, que são os sistemas de origem. Incluem fontes de dados com informações relevantes para a organização. Área de Extração, que envolve a extração de dados e transformações de preparação de dados. Área d Transformação, que envolve a limpeza, transformação e organização de dados extraídos. Área de Carregamento, onde os dados são carregados no Data Warehouse (DW). Metadados, que são informações descritivas de dados armazenados no DW. Ferramentas de Acesso e Consulta, que são ferramentas de Business Intelligence, que permitem ao utilizador aceder e consultar dados do DW. Camada de Apresentação, que é a interface com a qual o utilizador interage. Utilizadores Finais, que utilizam o DW.

Correção (Exercício 5):

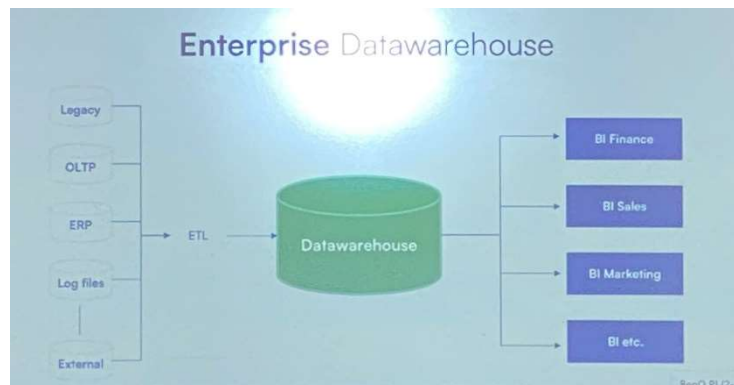


Figura 1 - Fotografia do Quadro

Exercício 6:

Alguns dos principais perigos de um Data Warehouse (DW) não integrado podem ser: Inconsistência de Dados, Dificuldades na Análise de Dados, Desperdício de Recursos Financeiros, ou seja, utilização de um excesso de recursos desnecessários, Dificuldades na Manutenção, Problemas de Segurança, Limitações de Escalabilidade, entre outros.

Correção (Exercício 6):

Exercício 7:

Alguns benefícios são a tomada de decisões informadas, através do acesso a dados consolidados e históricos dos mesmos; Consistência e

qualidade dos dados, e de acesso rápido; entre outros. Alguns obstáculos poderão ser a complexidade da integração de dados de diferentes fontes de dados; os Custos Financeiros elevados para a criação e manutenção de uma Data Warehouse; entre outros. Um desafio pode ser criar um mecanismo para ir buscar dados, e de uma forma mais responsiva.

Correção (Exercício 7):

Exercício 8:

Pode ser personalizada para ir buscar dados operacionais, de modo a satisfazer as necessidades específicas de uma organização.

Correção (Exercício 8):