***Business Intelligence***

**Datawarehouse** - segundo Inmon (1997), o DW é caracterizado como sendo uma coleção de dados orientados por assunto, integrada, não-volátil, variante no tempo e, o mais importante, fornece suporte à tomada de decisão de âmbito administrativo. Embora o conceito de Data Warehousing seja recente, ele se baseia em idéias que vinham sendo aplicadas em vários sistemas de informação há muitos anos (Inmon, 1997). Porém, há de se ressaltar que há uma distinção clara entre os bancos de dados tradicionais e os DWs, uma vez que os banco de dados tradicionais são transacionais (relacional, orientado a objetos, de rede, ou hierárquico), enquanto que os DW’s**têm a característica distinta de que são direcionados principalmente para aplicações de apoio às decisões. Assim, os DWs são otimizados para a recuperação de dados, e não para o processamento rotineiro de transações.** Complementando, os DWs proporcionam acesso aos dados para análise complexa, descoberta de conhecimento e tomada de decisão, sendo que eles fornecem suporte às demandas de alto desempenho de dados e informações de uma organização (ELMASRI e NAVATHE, 2005). **Características Datawarehouse**: orientado por assunto, integrado, não volátil, variável no tempo, que tem por objetivo dar suporte à **tomada de decisão.**

**Data mining** é uma expressão inglesa ligada à informática cuja tradução é mineração de dados. Consiste em uma funcionalidade que agrega e organiza dados, encontrando neles padrões, associações, mudanças e anomalias relevantes.

**Big data** se refere ao armazenamento de uma imensa quantidade de dados, bem como a capacidade de retirar valor dessas informações em velocidade rápida. Dessa forma, o **big data** se baseia em 5 V's. São eles: valor, volume, velocidade, variedade e veracidade

Classificação, regressão, classificação de séries temporais e clusterização são exemplos de tarefas de **Mineração de Dados** que podem ser implementadas por métodos de **Redes Neurais.**

Um***data* warehouse** é um repositório central de informações que podem ser analisadas para tomar decisões mais fundamentadas. Os dados fluem de sistemas transacionais, e de outras fontes para a data *warehouse*, normalmente com uma cadência regular, que **recebe informações de diferentes fontes, internas e externas.**

**Data Mart:** **É um pequeno data warehouse, abrangendo uma determinada área de assunto e oferecendo informações mais detalhadas sobre o mercado (ou departamento) em questão**.

Um Data Mart pode ser criado de duas maneiras:

1. Capturando dados diretamente de sistemas transacionais, cada Data Mart buscando as informações relevantes para o seu mercado;
2. Capturando dados de todos os sistemas transacionais em um Data Warehouse central, que por sua vez alimenta todos os Data Marts.

Data Marts: subconjuntos de um DW, repositório específico e volátil.

Traduzido como **Mercado de Dados**, trata-se de um subconjunto de dados de um **Data Warehouse**. Geralmente são dados referentes a um assunto em especial (Ex: Vendas, Estoque, Controladoria, entre outros) ou diferentes níveis de sumarização (Ex: Vendas Anual, Vendas Semestral, Vendas Mensal, entre outros), que focalizam uma ou mais áreas específicas. Seus dados são obtidos do DW – **indexados para suportar intensas pesquisas.**

Elepossibilita a análise de grandes volumes de dados, coletados dos sistemas transacionais **(OLTP)**. São as chamadas séries históricas que possibilitam uma melhor análise de eventos passados, oferecendo suporte às tomadas de decisões presentes e a previsão de eventos futuros. Por definição, os dados em um *DW***não são voláteis**, ou seja, eles não mudam, salvo quando é necessário fazer correções de dados previamente carregados. Os dados estão disponíveis somente para leitura e não podem ser alterados.

**1.Star Schema (Estrela):*Mais popular, mais rápido, com uma tabela por dimensão. É menos flexível e desperdiça espaço, pois repete as mesmas descrições ao longo de toda a tabela.***Possui uma tabela de fato e várias tabelas para cada dimensão, e propõe uma visão com a presença de dados redundantes nas tabelas de dimensão.

* **·Normalização: Não normaliza as dimensões.**
* **·​Espaço de armazenamento: Alto, pois por não haver normalização das tabelas, há muitos registros duplicados.**
* **·​Integridade: Mais propensa a problemas de integridade dos dados, devido a duplicação, dificultando futuras modificações e manutenções.**
* **·​Complexidade de query: Menos tabelas. Não há necessidade de relacionar muitas tabelas diferentes, logo a consulta é mais fácil.**
* ·**​Utilização: Data mart, pois o espaço de armazenamento exigido é pequeno e busca-se facilidade nas consultas de query.**

**2. Snow flake Schema (Floco de Neve):Variação do modelo estrela. Ele reduz o espaço de armazenamento, mais flexível, porém mais lento. *Apresenta o resultado da decomposição de uma ou mais dimensões que possuem hierarquias entre seus membros.***

* **·Normalização:Tabelas de dimensão são normalizadas.**
* **·​Espaço de armazenamento:Reduzido, pois a normalização reduz a quantidade de registros redundantes.**
* **·​Integridade:Maior facilidade para realizar futuras modificações e manutenções, pois há pouca redundância de dados.**
* **·​Complexidade de query:Por haver mais tabelas as consultas são maiores e mais complexas.**
* **·​Utilização:Data warehouse, pois dessa forma é possível economizar muito espaço de armazenamento.**

**Outra definição:**

* **Esquema multidimensional estrela:** todas as tabelas de dimensões se relacionam diretamente com tabelas de fato. Ou seja, uma tabela de dimensão não se relaciona com outra tabela de dimensão. **Não permite normalização**. Aumenta **desempenho**, mas não otimiza espaço em disco.
* **Esquema multidimensional flocos de neve: a**s tabelas de dimensões se relacionam com tabelas de fato, mas também se **relacionam entre elas**. Permite normalização até a **3ª FN**. Otimiza espaço, mas acaba diminuindo performance.

ETL: o ETL, do inglês Extract, Transform and Load, é o principal processo de condução dos dados até o armazenamento definitivo no DW. É responsável por todas as tarefas de extração, tratamento e limpeza dos dados, e inserção na base do DW.

**OLAP**: o OLAP, do inglês On-line Analytical Processing, na arquitetura de um DW se refere às ferramentas com capacidade de análise em múltiplas perspectivas das informações armazenadas.

**Data Mining**: Data Mining ou Mineração de Dados se refere às ferramentas com capacidade de descoberta de **conhecimento relevante dentro do DW**. Encontram correlações e padrões dentro dos dados armazenados.

Os 4 passos da modelagem dimensional são:

* Escolher o processo de negócio;
* Definir o grão;
* Identificar as dimensões;
* Identificar os fatos.

***Granularidade***: nada mais é do que o nível de detalhe ou de resumo dos dados existentes num banco de dados. Quanto maior for o nível de detalhes, menor será o nível de granularidade. O nível de granularidade afeta diretamente o volume de dados armazenados no banco de dados, e ao mesmo tempo o tipo de consulta que pode ser respondida.

Quando se tem um nível de granularidade muito alto o espaço em disco e o número de índices necessários, tornam-se bem menores, porém há uma correspondente diminuição da possibilidade de utilização dos dados para atender a consultas detalhadas.

Por exemplo, no caso de dados sobre vendas, o nível de granularidade muito baixo pode ser caracterizado pelo armazenamento de cada uma das vendas ocorridas para este produto, e um nível muito alto de granularidade seria o armazenamento dos somatórios das vendas ocorridas por mês.

***Operações OLAP***

* **Drill Through:** Ocorre quando o usuário passa de uma informação contida em uma dimensão para outra.
* **Slice:** Corta o cubo (extrai uma fatia), mas mantém a mesma perspectiva de visualização dos dados. Funciona como um filtro que restringe uma dimensão à apenas um ou alguns de seus valores.
* **Dice:** extrai um subcubo do cubo original executando uma operação de seleção em duas ou mais dimensões. Mudança de perspectiva da visão multidimensional, como se o cubo fosse girado. Permite descobrir comportamentos e tendências entre os valores das medidas analisadas em diversas perspectivas.
* **Drill Across:** O nível de análise dentro de uma mesma dimensão é alterado, ou seja, o usuário avança um nível intermediário dentro de uma mesma dimensão.
* **Pivot:** Adicionar ou rearranjar as dimensões das tabelas.
* **Drill Down:** proporciona uma visão mais detalhada de um conjunto de dados, descendo na hierarquia de uma dimensão.
* **Roll Up:** apresenta os dados cada vez mais agrupados ou sumarizados, subindo na hierarquia de uma dimensão.
* **Rotation:** permite visualizar dados de uma nova perspectiva.

**Oito passos da Data Mining:**

* **Associação: busca itens que ocorrem frequentemente de forma simultânea;**
* **Classificação: divide os atributos em tipos;**
* **Regressão: busca por uma função que ajude a mapear registros de um BD em um intervalo de valores numéricos reais;**
* **Agrupamento (Clusterização): segmenta os registros do conjunto de dados em subconjuntos (clusters). O objetivo é encontrar propriedades comuns de elementos;**
* **Sumarização: identifica e indica as semelhanças entre os registros;**
* **Detecção de Desvios: identifica registros que destoem o padrão no contexto de análise;**
* **Descoberta de sequências: identifica itens frequentes considerando um determinado período de tempo;**
* **Previsão de Séries Temporais: infere valores que a variável pode ou deve assumir no futuro;**

**O OLAP é uma interface com o usuário e não uma forma de armazenamento de dados**, porém se utiliza do armazenamento para poder apresentar as informações.

Os métodos de armazenamento são:

* **ROLAP (OLAP Relacional):** Os dados são armazenados de forma relacional.
* **MOLAP (OLAP Multidimensional):** Os dados são armazenados de forma multidimensional.
* **HOLAP (OLAP Híbrido):** Uma combinação dos métodos ROLAP e MOLAP.
* **DOLAP (OLAP Desktop):** O conjunto de dados multidimensionais deve ser criado no servidor e transferido para o desktop. Permite portabilidade aos usuários OLAP que não possuem acesso direto ao servidor.

Os métodos mais comuns de armazenamento de dados utilizados pelos sistemas OLAP são ROLAP e MOLAP, a única diferença entre eles é a tecnologia de banco de dados. O ROLAP usa a tecnologia RDBMS (Relational DataBase Management System), na qual os dados são armazenados em uma série de tabelas e colunas. Enquanto o MOLAP usa a tecnologia MDDB (MultiDimensional Database), onde os dados são armazenados em arrays multidimensionais.

Os dois fornecem uma base sólida para análise e apresentam tanto vantagens quanto desvantagens. Para se escolher entre os dois métodos devem-se levar em consideração os requisitos e a abrangência do aplicativo a ser desenvolvido.

**ROLAP é mais indicado para DATA WAREHOUSE pelo grande volume de dados**, a necessidade de um maior número de funções e diversas regras de negócio a serem aplicadas.

**MOLAP é mais indicado para DATA MARTS**, onde os dados são mais específicos e o aplicativo será direcionado na análise com dimensionalidade limitada e pouco detalhamento das informações.

Para se fazer uma comparação básica entre os dois métodos, as regras mais importantes são desempenho da consulta e desempenho do carregamento.

A senha é armazenada no **AD LDS** banco de dados em um objeto de usuário no atributo **unicodePwd**. Esse atributo pode ser escrito sob condições restritas, mas ele não pode ser lido. O atributo só pode ser modificado; ele não pode ser adicionado na criação do objeto ou consultado por uma pesquisa.