

Data Warehouse

Data Mining

Geraldo Robson, Mateus

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte - MG



Curso de Sistemas de Telecomunicações

O que é Data Warehouse?

Um Data Warehouse é um conjunto de dados baseado em assuntos, integrado, não volátil, e variável em relação ao tempo, de apoio às decisões gerenciais.

William Immon - 1992



Curso de Sistemas de Telecomunicações

Data Warehouse e Banco de Dados Operacional

OLAP (On-line Analytical Processing)

- Consultas dinâmicas - Dimensões
Dados históricos - exibem tendência de comportamento

Dados integrados

Número Reduzido de Transações
(1/dia, 1/semana)

Grande número de registros

Usuário: gerentes e executivos para tomada de decisões. Mas não há um padrão.

Analisa os dados

- OLTP (On-line Transaction Processing) - Consultas SQL prontas - Tabelas
- Dados operacionais
- Dados dispersos
- Grande número de Transações (frequentemente atualizado)
- Número de registros variável
- Usuário: profissional operacional da empresa e está preocupado em executar as operações de modo eficiente, e não como elas ocorrem
- Entrada de dados (armazenar)



Curso de Sistemas de Telecomunicações

O que é Data Warehousing?

Data Warehousing não é um produto, e sim uma estratégia que reconhece a necessidade de armazenar dados, separadamente, em sistemas de informação e consolidá-los, de forma a assistir diversos profissionais de uma empresa na tomada de decisões de modo rápido e eficaz.



Curso de Sistemas de Telecomunicações

Por que Data Warehousing?

A informação é o bem mais valioso para uma empresa;
Decisões precisam ser tomadas rapidamente e corretamente, usando todo dado disponível;

Usuários são "experts" em negócios, e não profissionais em tecnologia;

A quantidade de dados está dobrando a cada 18 meses, o que afeta o tempo de resposta e a habilidade de compreender seu conteúdo;

Estratégias de competição - diferencial positivo em relação a concorrência;

Análise do perfil do mercado;



Curso de Sistemas de Telecomunicações

Tipos de Data Warehouse

Marketing - Avalia a performance comercial de um produto ou serviço a partir de diversas perspectivas diferentes

Financeiro - Monitoram a performance comercial em termos financeiros

Comportamental - Contêm informações individuais a respeito de cada cliente e seus comportamentos.



Curso de Sistemas de Telecomunicações

Data Warehouse Incremental

Abordagens de implementação de DW (Top-Down e Up);

Projeto de impacto:

Projetos envolvem compromisso, recursos financeiros e pessoal, sujeitos a variações. (Tempo, orçamento, etc)
Custos envolvidos aumentam a pressão por resultados, recomenda-se iniciar um projeto-piloto de escopo e proporções reduzidos, mas com retorno considerável;

Criação de Data Marts Integrados (Data Marts extraem e ajustam porções de DWs aos requisitos específicos de grupos/departamentos) disponibilizados em prazos e custos menores, mais gerenciável.



Curso de Sistemas de Telecomunicações

Arquitetura de um Data Warehouse

Aquisição da Informação: Esta camada é responsável por reunir, refinar, limpar e agregar o dados dos sistemas de produção. O dado deve estar correto e prover um único significado para facilitar na tomada de decisão.

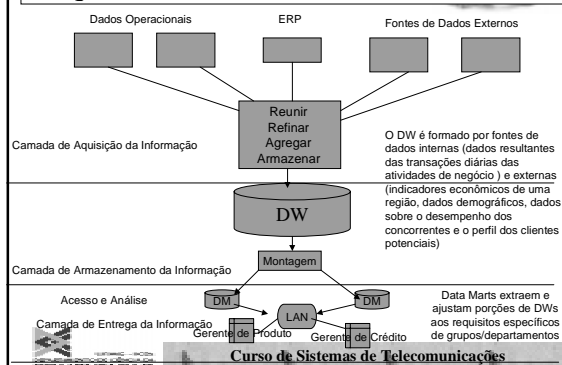
Armazenamento da Informação: Esta camada provê o armazém da informação.

Disponibilização da Informação: Esta camada é parte da arquitetura que suportará um conjunto de ferramentas de apresentação e análise. (Relatórios e consultas)

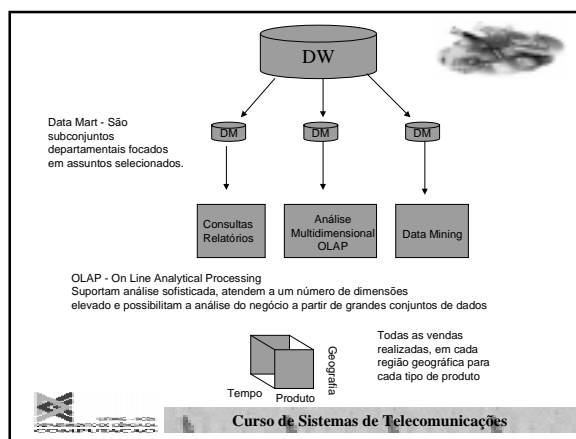


Curso de Sistemas de Telecomunicações

Arquitetura de um Data Warehouse (DW)



Curso de Sistemas de Telecomunicações



Curso de Sistemas de Telecomunicações

O que é DW para Intranet?

Combinação de tecnologias que permite que seus usuários gerem dinamicamente uma pesquisa no banco de dados, analisando e formatando os resultados na forma de arquivos texto ou imagem.



Curso de Sistemas de Telecomunicações

Por que DW para Intranet?

Melhora a capacidade de comunicação e colaboração dentro da empresa;

Atinge um número grande de usuários sem conhecimentos técnicos.



Curso de Sistemas de Telecomunicações

O que é Data Mining?

Mineração dos dados : usa complexas pesquisas em grandes volumes de dados para explorar e identificar relacionamentos entre variáveis que antes eram previamente independentes. Isso permite que as organizações gerenciem relacionamentos de causa e efeito mais eficientemente, e antecipem políticas para isso.



Curso de Sistemas de Telecomunicações

Diferenças entre DM e OLAP

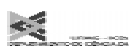
Data Mining

Método alternativo (trabalha contra todos os dados de um grupo) e automático (estuda o dados e apresenta o “achado”) para descobrir padrões nos dados. Questão desconhecida e a ferramenta é utilizada para busca de conhecimentos.

OLAP

Trabalha com alguns dados e tem que perfurar (drill-down) em busca de detalhes. Explorados manualmente.

Exploração através da verificação - o analista elabora uma hipótese para confirmá-la ou refutá-la.



Curso de Sistemas de Telecomunicações

Como implementar um Data Mining

Algoritmos baseados em redes neurais, estatísticas ou algoritmos de aprendizado, capazes de explorar gigantescas bases de dados em busca de informações que ajudam as empresas a decifrar o comportamento do consumidor, a compra de suprimentos, administrar as áreas comerciais e financeira de forma transparente.

Ferramentas capazes de aprender a partir dos dados, gerando e validando enormes volumes de hipóteses e destacando conhecimento interessante para o usuário, descobrindo conhecimento novo, útil e interessante.



Curso de Sistemas de Telecomunicações

Aplicações de Data Mining

Cross-selling: Identificar associação entre produtos, procurando tirar proveito dessas associações.

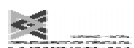
Up-selling: Identificar entre os clientes atuais aqueles que tem potencial para adquirir produtos que proporcionem maior lucro para a companhia. Em um serviço de TV por assinatura, por exemplo, uma forma de realizar *up-selling* seria identificar o perfil dos compradores de *Pay-Per-View* e utilizar este perfil para dirigir uma campanha aos clientes que tem as mesmas características, mas ainda não compram este produto;



Curso de Sistemas de Telecomunicações

Aplicações de Data Mining

Fidelização: Descobrir o perfil, situação e combinação de fatores que estão relacionados com a perda de clientes, para que se trabalhe estes clientes e mantenha o relacionamento, dentro da filosofia de que manter um cliente conquistado é mais barato que conquistar novos clientes.




Curso de Sistemas de Telecomunicações

Algumas barreiras do uso de DM

- Altos custos;
- Complexidade de ferramentas;
- Preparação dos dados para mineração (80% do trabalho). Os dados devem estar limpos, consistentes e livres de excessivas nulidades;
- Dificuldade em se estimar a taxa de retorno do investimento. Uma vez que o objetivo do DM é descobrir tendências em dados invisíveis de outra maneira, torna-se impossível estimar a partir de algo desconhecido.



Curso de Sistemas de Telecomunicações




Um projeto de Data Warehouse nunca termina! As fontes de dados mudam, as necessidades dos usuários também mudam, o número deles aumenta, as regras de mercado se alteram, os usuários demandarão mais informações. Enfim, data warehousing não é um destino; é uma viagem.



Curso de Sistemas de Telecomunicações

Bibliografia



Data Warehouse - Conceitos e Soluções
Adelize Generini de Oliveira
Data Mining - A hands-on approach for bussiness professionals
Robert Groth
Data Warehouse Intranet
Internet
<http://www.datawarehouse.inf.br/>



Curso de Sistemas de Telecomunicações