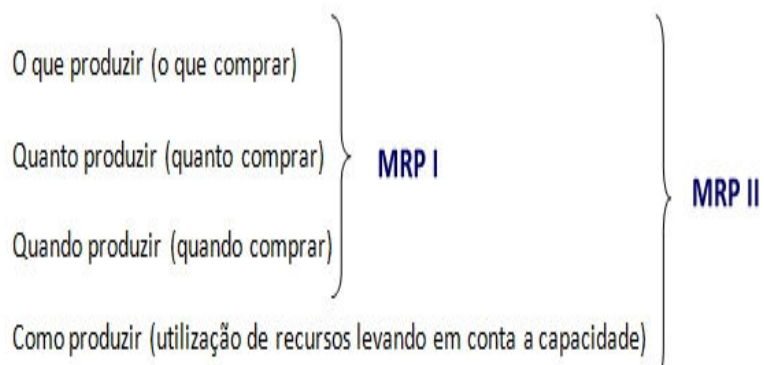


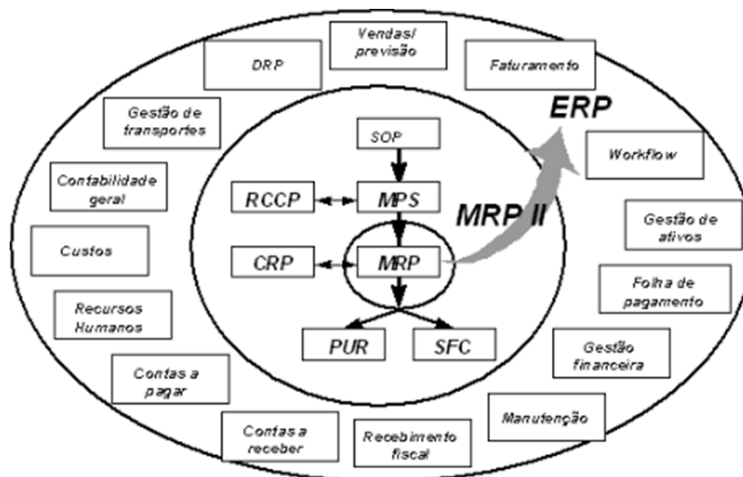
FUNDAMENTOS DE SISTEMAS DE INFORMAÇÃO

Rafael D. Ribeiro, M.Sc,PMP.
rafaeldiasribeiro@gmail.com
<http://www.rafaeldiasribeiro.com.br>

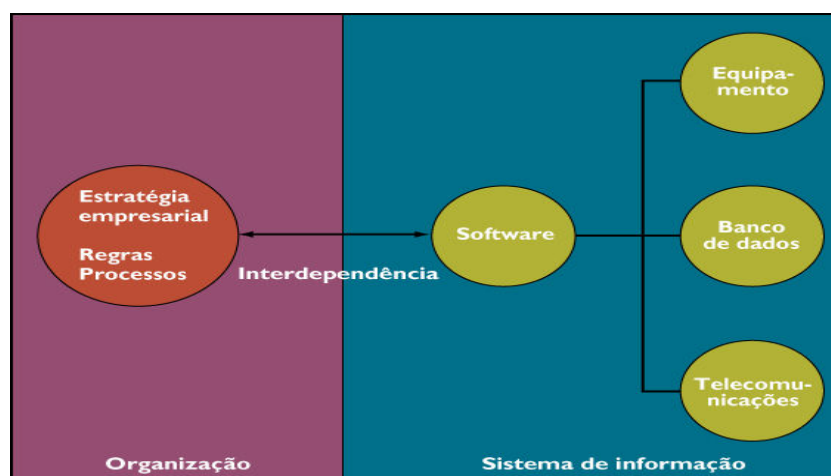
Lembrando... Aula 4



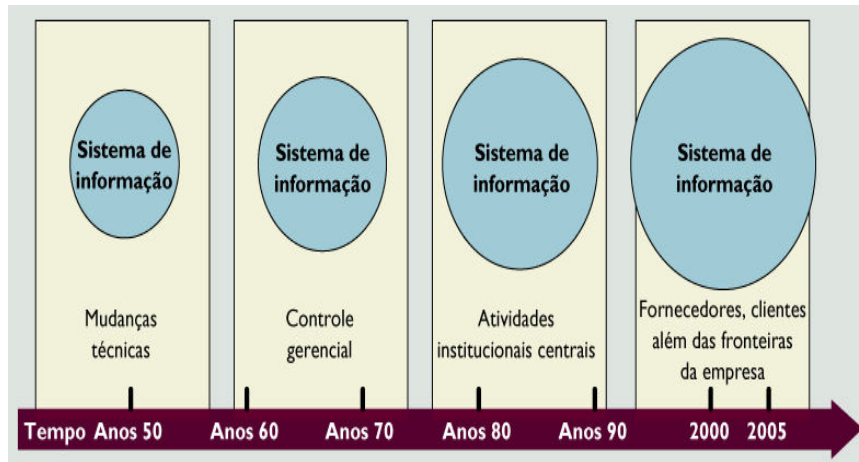
Lembrando... Aula 4



Sistemas de apoio às decisões

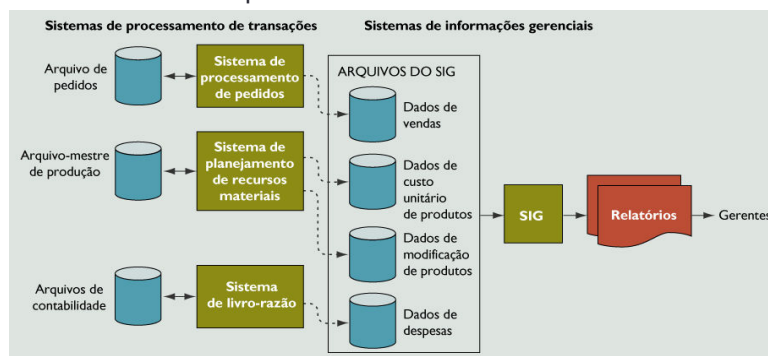


Sistemas de apoio às decisões

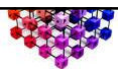


- Os Sistemas de Informação Gerencial (SIG) devem ser capazes de captar, processar, armazenar e distribuir dados e informações para quem deles necessitam, especialmente as pessoas envolvidas no processo decisório, visando o planejamento, a execução e o controle das atividades, bem como a tomada de decisão nas organizações.
- Á nível gerencial, a entrada tem alto volume de dados, o processamento tem modelos simples, as saídas são representadas por relatórios/sumários executivos e ou usuários são os gerentes de nível médio.

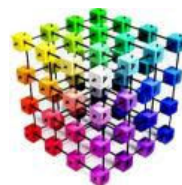
- Os SIG têm decisões estruturadas e semi-estruturadas, são orientados para o controle de relatórios, utiliza dados passados e presentes, tem orientação interna e o processo de projeto demorado.
- A sua estrutura é dada por:



- Hoje em dia uma organização precisa utilizar toda informação disponível para criar e manter vantagem competitiva.
- Sai na frente a organização que consegue tomar decisões corretas e rápidas. Com esta importante tarefa nas mãos, profissionais tomadores de decisão tais como executivos, gerentes e analistas, exigem dos sistemas de suporte à decisão (*Decision Support Systems - DSS*) mais recursos para análise, front-ends que suportem consultas ad hoc, interfaces gráficas apropriadas, etc.



- A idéia de data warehouse é integrar os dados internos e externos de uma organização em uma estrutura única permitindo uma melhor utilização dos dados pelos analistas, gerentes e executivos.
- Uma vez obtida a integração, sistemas como OLAP (On-Line Analytical Processing) e data mining fornecem mecanismos sofisticados para análise dos dados.



Data Warehouse



- Um data warehouse é uma coleção de dados orientada por assuntos, integrada, variante no tempo, que tem por objetivo dar suporte aos processos de tomada de decisão. (William H.Inmon)
- Ele é um ambiente expansível e planejado para a análise de dados não voláteis.
- Estes dados são logicamente e fisicamente transformados, provenientes de múltiplas aplicações, atualizados e mantidos por um longo período de tempo, expressos em termos do negócio e resumidos para uma análise eficiente. Assim, um Data Warehouse é uma plataforma com dados integrados e qualidade melhorada para apoiar vários DSS (Sistema de Apoio de Decisão), aplicações de EIS (Sistema de Informação Executiva) e processos empresariais.

Data Warehouse



- O data warehouse é um banco de dados contendo dados extraídos do ambiente de produção da empresa, que foram selecionados e depurados, tendo sido otimizados para processamento de consulta e não para processamento de transações.
- Em geral, um data warehouse requer a consolidação de outros recursos de dados além dos armazenados em base de dados relacionais, incluindo informações provenientes de planilhas eletrônicas, documentos textuais, etc.

Data Warehouse



- O objetivo de um data warehouse é fornecer uma "imagem única da realidade do negócio" (Richard Hackathorn).
- De uma forma geral, sistemas de data warehouse compreendem um conjunto de programas que extraem dados do ambiente de dados operacionais da empresa, um banco de dados que os mantém, e sistemas que fornecem estes dados aos seus usuários.

Data Warehouse

Por que investir em um ...



- As decisões precisam ser tomadas rapidamente e corretamente, usando todos os dados disponíveis;
- Os usuários de sistemas de informações são especialistas de domínio de negócio, não profissionais de computação;
- O volume de dados dobra a cada 18 meses, o que afeta o tempo de resposta e incontestavelmente a habilidade em compreender seu conteúdo.
- A competição aumenta dia após dia nas áreas de inteligência empresarial, bem como o valor agregado de informações.

Data Warehouse

Por que investir em um ...



- Os dados armazenados para análise do negócio devem ser separados dos dados do sistema operacional. Muitas razões para esta separação têm sido levantada no decorrer dos anos, a saber:
 - a performance do sistema de informação;
 - a garantia da qualidade da informação obtida - completa e correta;
 - a não perturbação dos sistemas operacionais através de requisições de consultas pesadas, etc.
- Finalmente, os avanços tecnológicos e as mudanças na natureza dos negócios tornam a análise dos mesmos muito mais complexa e sofisticada.

Data Warehouse

De acordo com Bill Inmon, autor de Building the Data Warehouse e guru que é considerado o originador do conceito de data warehousing, existem basicamente quatro características que descrevem um data warehouse:

- **orientado ao assunto:** os dados são organizados de acordo com o assunto ao invés da aplicação.
 - Por exemplo, um companhia de seguros utilizando um data warehouse poderia organizar seus dados por consumidor, prêmio e franquia, ao invés de diferentes produtos (automóvel, vida, etc.). Os dados são organizados por assunto, contendo somente as informações necessárias para o processo de suporte à decisão.

Data Warehouse

- **Integrado:** quando os dados residem em várias aplicações diferentes no ambiente operacional, a codificação dos dados normalmente é inconsistente.
 - Por exemplo, em uma aplicação o gênero pode estar codificado como "m" e "f" e em uma outra, como 0 e 1. Quando os dados são movidos de um ambiente operacional para o data warehouse, estes assumem uma convenção de código consistente - neste caso, o campo gênero é transformado em "m" e "f".
- **variação em relação ao tempo:**
 - O data warehouse contem um lugar para o armazenamento de dados que têm 5, 10 anos ou mais, para serem utilizados em comparações, tendências e previsões. Estes dados não são atualizados.

Data Warehouse

- **Não-volatilidade:**
 - Os dados não são atualizados ou modificados em nenhum momento uma vez que estes entram no data warehouse, são somente acessados.

OLAP (On-Line Analytical Processing)

- OLAP é uma categoria de tecnologia de software que habilita analistas, gerentes e executivos ter acesso rápido, consistente e interativo a uma grande variedade de possíveis visões de uma informação.
- O OLAP aplica-se a todas as funções analíticas requeridas para a criação de informações úteis a partir dos dados armazenados em um DW, consiste na geração de consultas ao banco de dados, execução de cálculos e formatação e permite que os usuários executem funções de análise desses dados através das dimensões do DW.

OLAP (On-Line Analytical Processing)

- A funcionalidade do OLAP é caracterizada pela análise dinâmica multidimensional de dados consolidados, incluindo *drill-down* para "penetrar" os níveis de consolidação.

- *Drill-down* é um método de exploração detalhada de dados que é utilizado na criação de níveis de sumarização de dados. Os níveis de *drill-down* dependem da granularidade dos dados no Data Warehouse.
- A implementação de um sistema de OLAP é multi-usuário e cliente/servidor.

Data Mining

- Data Mining consiste em um processo analítico projetado para explorar grandes quantidades de dados (tipicamente relacionados a negócios, mercado ou pesquisas científicas), na busca de padrões consistentes e/ou relacionamentos sistemáticos entre variáveis e, então, validá-los aplicando os padrões detectados a novos subconjuntos de dados.



Data Mining

- Data Mining é uma das fases mais importantes em um processo KDD (Knowledge Discovery in Databases) ou seja, metodologia DCBD (Descoberta de Conhecimento em Bases de Dados), e tem como finalidade principal a busca de padrões de negócio.
- O processo KDD é um sistema que pode extrair o conhecimento de múltiplas bases, porém nem todos os dados que compõem as bases servem para um sistema descobrir conhecimento.



Data Mining - Fases

- **Pré-processamento dos dados:** o sucesso dos resultados obtidos no processo de KDD depende diretamente da qualidade dos dados de entrada;
- **Redução de dado:** é uma forma de pré-processamento que visa obter uma representação reduzida dos dados, mas que produz os mesmos resultados analíticos;
- **Preparação dos dados:** as três fases iniciais do KDD (seleção, pré-processamento transformação), exigem bastante tempo, aproximadamente entre 60 e 80% do tempo utilizado em todo o processo, sendo que a maior parte desse tempo é consumida com a limpeza dos dados. O foco será dado no pré-processamento, no *Data Mining* e na Interpretação/Avaliação do Conhecimento.

Data Mining - Fases

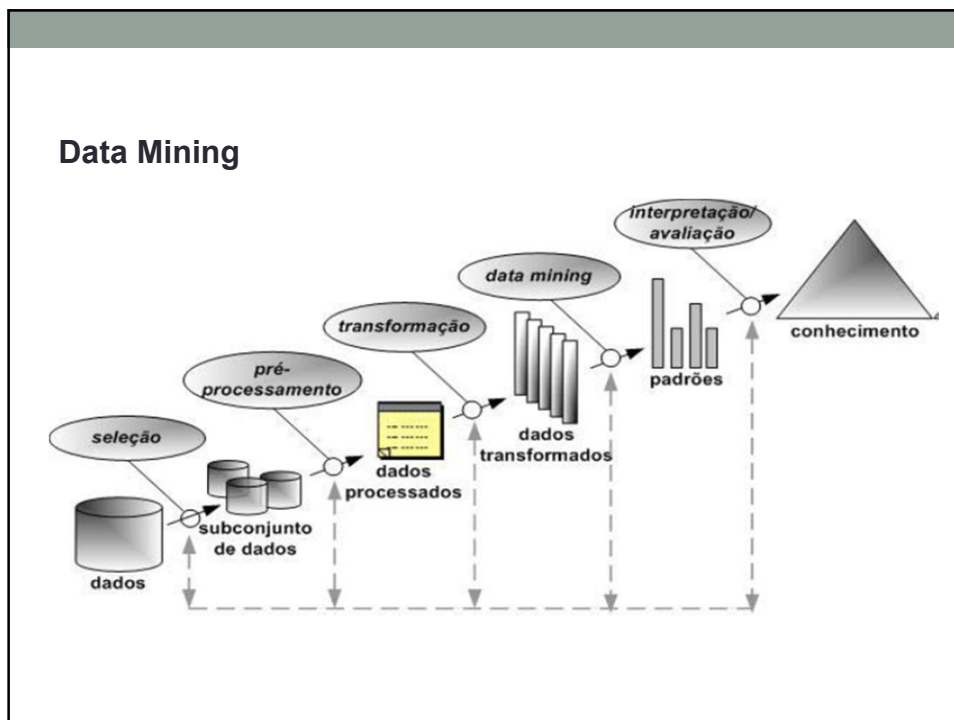
- **Limpeza dos dados:** a verificação da consistência das informações, a correção de possíveis erros e o preenchimento ou a eliminação de valores nulos e redundantes. São identificados e removidos os dados, duplicados e/ou corrompidos. A execução dessa fase corrige a base de dados eliminando consultas desnecessárias que seriam executadas pelo **algoritmo** minerador e que afetariam o seu processamento
- **Transformação de Dados ou Codificação dos Dados:** tem como objetivo principal converter o conjunto bruto de dados em uma forma padrão de uso;
- **Redução de dados:** é uma forma de pré-processamento que visa obter uma representação reduzida dos dados, mas que produz os mesmos resultados analíticos;

Data Mining - Fases

- **Mineração de dados:** Caracteriza-se pela existência do algoritmo minerador (*Data Mining*), que diante da tarefa especificada será capaz de extrair de modo eficiente conhecimento implícito e útil de um banco de dados Usa técnicas baseadas em descobertas por meio de procura de padrões dos dados, o que é feito com o emprego de algoritmos inteligentes para encontrar relações fundamentais entre os dados.

Permitem avaliar como as perguntas se relacionam com as respostas (padrões e relações) encontradas, fornecendo uma base de regras que servem de apoio aos processos de tomada de decisão.

- **Interpretação do conhecimento descoberto, ou algum processamento desse conhecimento:** deve ser incluída no algoritmo minerador, ou implementado separadamente;



Data Mining

- A premissa do Data Mining é uma argumentação ativa, isto é, em vez do usuário definir o problema, selecionar os dados e as ferramentas para analisar tais dados, as ferramentas do Data Mining pesquisam automaticamente os mesmos a procura de anomalias e possíveis relacionamentos, identificando assim problemas que não tinham sido identificados pelo usuário.
- As ferramentas de Data Mining analisam os dados, descobrem problemas ou oportunidades escondidas nos relacionamentos dos dados, e então diagnosticam o comportamento dos negócios, requerendo a mínima intervenção do usuário. Assim, ele se dedicará somente a ir em busca do conhecimento e produzir mais vantagens competitivas.

Tipo de ferramenta	Questão básica	Exemplo de resposta	Usuário típico e suas necessidades
Pesquisa e Relatórios	"O que aconteceu?"	Relatórios mensais de vendas, histórico do inventário	Dados históricos, habilidade técnica limitada
OLAP	"O que aconteceu e por que?"	Vendas mensais versus mudança de preço dos competidores	Visões estáticas da informação para uma visão multidimensional; tecnicamente astuto
Sistema de Informação Executivo (EIS)	"O que eu preciso saber agora?"	Memorandos, centros de comando	Informações de alto nível ou resumidas; pode não ser tecnicamente astuto
Data Mining	"O que é Interessante?" "O que pode acontecer?"	Modelos de previsão	Tendências e relações obscuras entre os dados; tecnicamente astuto

BIBLIOGRAFIA

Albertin, Alberto Luiz. *Estratégia de Governança de Tecnologia de Informação: Estrutura e Práticas* / Alberto Luiz Albertin, Rosa Maria de Moura Albertin. Rio de Janeiro. Elsevier, 2010.

Turban, Efrain. *Administração da Tecnologia da Informação: Teoria e Prática* / Efrain Turban, R. Kelly Rainer, Richard E. Potter; tradução Daniel Vieira. Rio de Janeiro:Elsevier,2005. 6ª reimpressão.