

[MAC0313]

Introdução aos Sistemas de Bancos de Dados

Aula 1

Modelagem e Implementação de Bancos de Dados (Conceitos Introdutórios)

Kelly Rosa Braghetto
(kellyrb@ime.usp.br)

DCC-IME-USP

06 de agosto de 2014

Qual é a relação entre Estatística e Bancos de Dados?

- ▶ Segundo a Wikipédia ¹:

*A estatística é uma ciência que se dedica à coleta, análise e interpretação de **dados**.*

- ▶ Ferramentas de software estatísticas (também chamadas de “pacotes estatísticos”) focam a análise dos dados
- ▶ Sistemas de bancos de dados focam o armazenamento e a manipulação dos dados, levando em conta sua estrutura
 - ▶ Com isso, eles amparam a coleta e interpretação dos dados

¹<http://pt.wikipedia.org/wiki/Estatística>)

Sobre os pacotes estatísticos ...

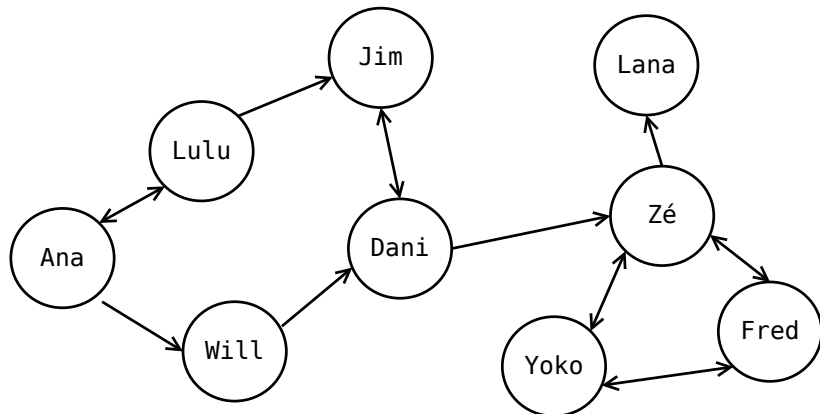
- ▶ Geralmente disponibilizam procedimentos que recebem com parâmetro de entrada vetores (*arrays*) de dados e que geram saídas na forma de gráficos, tabelas ou outros vetores de dados

Deficiências comuns desse tipo de software:

- ▶ Não manter os dados junto de suas respectivas informações estruturais
- ▶ Não garantir que a análise estatística realizada será apropriada para os dados fornecidos como entrada
- ▶ Ser susceptível a dados de má qualidade, como dados incompletos ou inconsistentes

Um exemplo motivacional :)

Dados de uma “rede social”



Um exemplo motivacional :)

Dados de uma “rede social”

- ▶ Queremos responder perguntas como:
 - ▶ “Quem conhece o Will?”
 - ▶ “Existe alguém que não é conhecido por ninguém?”
 - ▶ “Por quantas pessoas uma pessoa é conhecida em média?”
- ▶ A representação baseada em grafo facilita o armazenamento dos dados, a realização de consultas e a análise estatística?

Um exemplo motivacional :)

Uma outra representação da rede social

Jim	-	Dani
Lulu	-	Jim, Ana
Lana	-	
Ana	-	Lulu, Will
Dani	-	Jim, Zé
Zé	-	Lois, Yoko, Fred
Will	-	Ana, Dani
Yoko	-	Zé, Fred
Fred	-	Zé, Yoko

Essa representação “cabe” em um vetor retangular (= tabela) com 4 colunas, apropriado para manipulação em um pacote estatístico

Um exemplo motivacional :)

Problemas com a representação tabular proposta

- ▶ Responder à pergunta “Quem conhece o Jim?” continua difícil
- ▶ A modificação dos dados pode gerar inconsistências.

Exemplos:

- ▶ Se o nome de uma pessoa foi escrito errado, a correção do erro exigirá a modificação de todas as ocorrências dele na tabela
- ▶ Se houver duas pessoas que possuam um mesmo nome no grupo, isso gerará confusão
- ▶ Se uma pessoa que já possui três conhecidos conhecer alguém novo, não será possível incluir esse dado na rede
- ▶ Quando novas pessoas são incluídas, é preciso verificar se todas as pessoas que elas conhecem também estão incluídas como registros separados

Solução: usar sistemas de bancos de dados

O que é um **banco de dados**?

- ▶ **Banco de dados** – coleção de dados relacionados
- ▶ **Dados** – fatos conhecidos que podem ser registrados e que possuem significado implícito

Problema: essa definição é genérica demais!

Propriedades implícitas de um BD

1. Representar (geralmente!) algum aspecto do mundo real = **minimundo** ou **UoD** (Universo de Discurso)
As mudanças no minimundo são refletidas no BD
2. Ser uma **coleção lógica e coerente de dados** com algum **significado inerente**
Uma coleção “aleatória” de dados não é um BD!
3. Ser projetado, construído e povoado com dados que possuem um objetivo específico
Um BD deve possuir um grupo provável de usuários e algumas aplicações pré-concebidas, nas quais esses usuários estão interessados

Bancos de dados

⇒ **Resumindo**: um BD possui alguma **fonte** (de onde os dados são derivados), algum grau de **interação com eventos do mundo real** e um **público** que está ativamente interessado no conteúdo do BD

Outras características:

- ▶ BDs têm complexidade e tamanho variáveis
- ▶ BDs podem ser informatizados ou mantidos manualmente

Um BD informatizado pode ser criado e mantido por:

- ▶ um grupo de programas de aplicação (criados especificamente para essa tarefa)
- ▶ um **sistema de gerenciamento de banco de dados (SGBD)**

Exemplo da dimensão que um BD pode assumir

Amazon.com (em 2007)

- ▶ dados de mais de 20 milhões de livros, CDs, DVDs, jogos, eletrônicos, etc.
- ▶ o banco de dados ocupa mais de 2 terabytes (2 TB), armazenados em centenas de servidores
 - ▶ 1 terabyte = 1.000 gigabytes = 1.000.000.000.000 bytes
- ▶ cerca de 15 milhões de visitantes acessam a Amazon.com todos os dias
- ▶ são necessários mais de 100 funcionários para manter o BD da Amazon atualizado

Exemplo da dimensão que um BD pode assumir

- ▶ **Facebook (abr/2009)**: cerca de 2,5 petabytes de dados de usuários + 15 terabytes/dia
 - ▶ 1 petabyte = 1.000 terabytes = 1 quadrilhão de bytes (210.000 DVDs)
- ▶ **eBay (mai/2009)**: cerca de 6,5 petabytes de dados dos usuários + 50 terabytes/dia
- ▶ **Large Hadron Collider (LHC)**: produz cerca de 25 petabytes/ano, processa cerca de 1 petabyte de dados todos os dias e já possui mais de 100 petabytes armazenados desde o início do projeto (700 anos de filmes armazenados com qualidade full HD)

Sistema de Gerenciamento de Banco de Dados (SGBD)

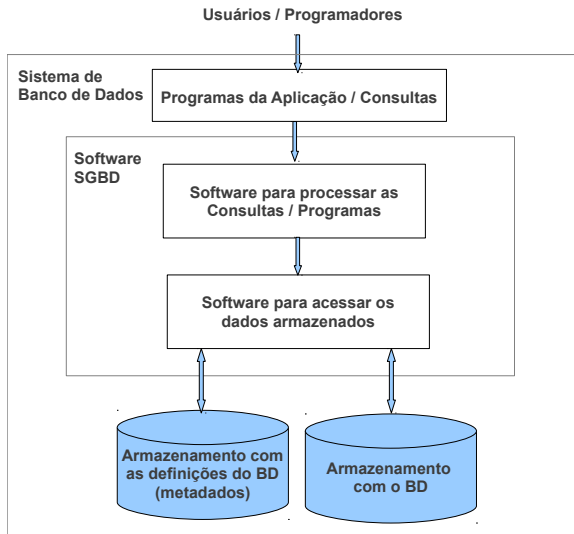
É um sistema de software de propósito geral que facilita o processo de **definição, construção, manipulação e compartilhamento** de BDs entre vários usuários e aplicações.

- ▶ **Definir um BD** $\Rightarrow$ especificar os tipos, as estruturas e as restrições para os dados que serão armazenados no BD
- ▶ **Construir um BD** $\Rightarrow$ armazenar os dados em algum meio de armazenamento (controlado pelo SGBD)
- ▶ **Manipular um BD** $\Rightarrow$ realizar funções como consultas ao BD para recuperar dados específicos, atualizar o BD para refletir mudanças no mini-mundo, etc.
- ▶ **Compartilhar um BD** $\Rightarrow$ permitir que múltiplos usuários e programas acessem-no simultaneamente

Outras funções importantes de um SGBD

- ▶ **Proteger os dados** – contra falhas de hardware ou software e contra acessos não autorizados ou maliciosos
- ▶ **Manter os dados por um longo período de tempo** – permitindo que o sistema evolua acompanhando as mudanças dos requisitos ao longo do tempo

Sistema de Banco de Dados = Banco de Dados + SGBD



O acesso a um BD

Um **programa de aplicação** é um programa que acessa um banco de dados enviando **consultas** ou **transações** para o SGBD.

- ▶ **Consulta** – comando que recupera dados do BD
- ▶ **Transação** – comando que lê ou escreve dados do/no BD

Banco de dados × processamento de arquivos tradicional

- ▶ No **processamento de arquivos tradicional**: cada usuário define e implementa os arquivos necessários para uma aplicação específica (como parte da aplicação)

Problema: possível redundância na definição e armazenamento de dados ⇒ desperdício de espaço e trabalho redundante na manutenção de dados comuns a mais de uma aplicação

- ▶ Nos **bancos de dados**: um único repositório de dados é mantido

Uma vez definido, o repositório passa a ser acessado por diversos usuários e aplicações

Outras características importantes dos bancos de dados

- ▶ Natureza auto-descritiva
- ▶ Isolamento entre programas e dados, e abstração de dados
- ▶ Suporte a visões múltiplas dos dados
- ▶ Compartilhamento de dados e processamento de transações multiusuários

Vantagens do uso de um SGBD

1. Controle de redundâncias
2. Restrição do acesso não autorizado
3. Armazenamento persistente para objetos de programas e estruturas de dados
4. Estruturas de armazenamento e técnicas de busca para o processamento eficiente de consultas
5. Mecanismos de backup e recuperação
6. Múltiplas interfaces de usuário
7. Capacidade de representação de relacionamentos complexos entre dados
8. Imposição de restrições de integridade
9. Possibilidade de deduzir dados e executar ações por meio de regras

Vantagens do uso de um SGBD

Implicações adicionais:

1. Potencial para garantir padrões
2. Redução no tempo de desenvolvimento de aplicações
3. Flexibilidade
4. Disponibilidade de informações atualizadas
5. Economias de escala

Analizando a viabilidade do uso de um SGBD

Os custos (de tempo e dinheiro) envolvidos no uso de um SGBD se relacionam a:

1. Os altos investimentos iniciais em hardware, software e treinamento
2. A generalidade que o SGBD fornece para a definição e o processamento de dados
3. O esforço adicional necessário para prover segurança, controle de concorrência, recuperação e integridade dos dados

Quando é melhor **não** usar um SGBD

O uso de arquivos convencionais é mais aconselhado que o uso de SGBDs nas seguintes situações:

1. O BD e suas aplicações são simples, bem definidos e sem previsão de mudanças
2. A sobrecarga do SGBD pode impedir que requisitos de desempenho (como em programas de tempo-real) não sejam atendidos
3. O acesso de múltiplos usuários aos dados não é necessário

Referências Bibliográficas

- ▶ *Sistemas de Bancos de Dados* (6ª edição), Elmasri e Navathe. Pearson, 2010.
Capítulo 1
- ▶ *A First Course in Database Systems* (3ª edição), Ullman e Widom. Prentice Hall, 2007.
Capítulo 1
- ▶ *Introdução a Sistemas de Bancos de Dados* (8ª edição), Date. Campus, 2004.
Capítulos 1