

Banco de Dados

Aula 1 - Introdução a Banco de Dados

Introdução

Um *Sistema Gerenciador de Banco de Dados* (SGBD) é constituído por um conjunto de dados associados a um conjunto de programas para acesso a esses dados.

O principal objetivo de um SGBD é proporcionar um ambiente tanto *conveniente* quanto a *eficiente* para a recuperação e armazenamento das informações do banco de dados.

Sistemas de banco de dados são projetados para gerir grandes volumes de informações. O gerenciamento de informações implica a definição das **estruturas de armazenamento das informações** e da definição dos **mecanismos para a manipulação dessas informações**. Também um sistema de banco de dados deve **garantir a segurança** das informações armazenadas contra eventuais problemas com o sistema, além de impedir tentativas de acesso não autorizadas. Também se os dados são compartilhados por diversos usuários o sistema deve evitar a ocorrência de resultados anômalos.

Por que Sistemas de Banco de Dados

Um **Sistema de processamento de arquivos** convencional onde, registros permanentes são armazenados em vários arquivos e diversos programas de aplicação são escritos para extrair e gravar registros nos arquivos apropriados, podem ser apropriadamente gerenciados pelos sistemas operacionais existentes. Mas, estes sistemas apresentam numerosas desvantagens:

- **Inconsistência e Redundância de Dados:** Como arquivos e programas normalmente são criados e mantidos por diferentes programadores, em geral é comum que os arquivos possuam formatos diferentes e os programas estejam escritos em diferentes linguagens de programação. Além disso, a mesma informação pode estar repetida em mais de um arquivo.
- **Dificuldade de Acesso aos Dados:** Normalmente em um sistema de processamento de arquivos comum, quando existe uma necessidade de uma relação de empregados, por exemplo, segundo uma determinada condição, se esta situação não tiver sido prevista inicialmente no sistema, é necessário que um novo programa seja desenvolvido e que gere esta lista solicitada. O fato é que este ambiente não está preparado para atender 'as necessidades de recuperação de informações de modo eficiente.

•**Isolamento dos Dados:** Como os dados estão dispersos em vários arquivos, e estes arquivos podem apresentar diferentes formatos, é difícil escrever novas aplicações para a recuperação apropriada destes dados.

•**Problemas com Integridade:** Os valores dos dados atribuídos e armazenados em um banco de dados devem satisfazer certas restrições para manutenção da consistência. O problema aumenta quando as restrições atingem diversos itens de dados em diferentes arquivos.

•**Problemas de atomicidade:** Um sistema computacional está sujeito a falhas. É imprescindível garantir que, uma vez detectada uma falha, os dados sejam salvos em seu último estado consistente, anterior a ela. Por exemplo, uma operação de transferência bancária entre contas correntes, deve ser uma operação atômica, ou seja, deve ocorrer por completo, ou não ocorrer.

•**Anomalias no acesso concorrente:** Muitos sistemas permitem atualizações simultâneas dos dados para aumento do desempenho do sistema como um todo e para melhores tempos de resposta. Este tipo de interação pode resultar em inconsistência de dados. Por exemplo, dois saques simultâneos a uma mesma conta corrente.

•**Problemas de Segurança:** Nem todos os usuários de banco de dados estão autorizados ao acesso a todos os dados. Em um sistema bancário, por exemplo, os funcionários do departamento de pessoal não deveriam ter acesso às informações dos clientes do banco, só apenas ao conjunto de pessoas ao qual o seu departamento lhe diz respeito.

Estas e outras dificuldades provocaram o desenvolvimento dos SGBDs. Iniciaremos o estudo dos sistemas de bancos de dados, mostrando, para isso, conceitos e algoritmos que foram desenvolvidos para a efetiva manipulação dos mesmos.

Considerações Básicas, Abstração e Modelos de Dados

Bancos de Dados e Sistemas de Banco de Dados têm se tornado um componente essencial todos os dias da vida moderna. Aplicações de banco de dados tradicionais dizem respeito a sistemas bancários, sistemas de reservas aéreas, sistemas de controle de estoque.

Com o avanço da tecnologia, novos tipos de bancos de dados foram desenvolvidos, com aplicações e uso específicos. **Bancos de Dados Multimídia** podem, por exemplo, armazenar figuras, vídeo-clips e mensagens de sons. **Bancos de Dados Geográficos** podem armazenar e analisar mapas, dados climáticos e imagens de satélite. **DataWarehouses** e **Processamentos Analítico On-line** (OLAP) são sistemas usados em muitas empresas para extrair e analisar informações úteis de sistemas de banco de dados muito grandes para a tomada de decisões.

Tecnologias de Tempo Real e de **Banco de Dados Ativos** são usados no controle industrial e em processo de manufatura. Técnicas de busca em bancos de dados, com aplicações na Web, têm melhorado a eficiência dos mesmos. No entanto, temos que entender o básico das aplicações de banco de dados tradicionais.

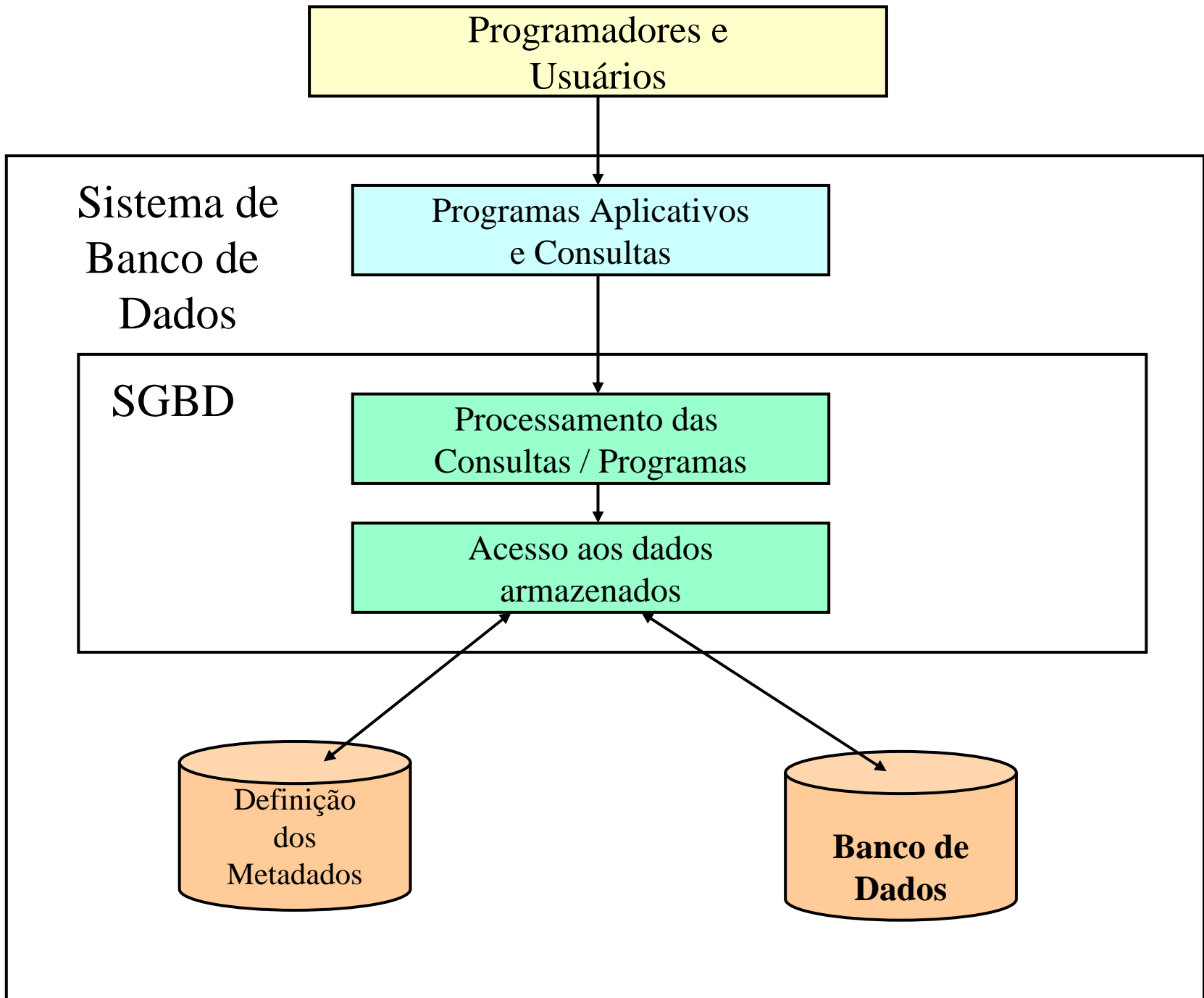
Um banco de dados pode ser considerado como uma coleção de dados relacionados. Por dado entendem-se fatos conhecidos que podem ser armazenados e que possuem um significado implícito. Um banco de dados possui as seguintes propriedades implícitas:

- Representa alguns aspectos do mundo real, muitas vezes chamado de mini-mundo ou universo do discurso.
- É uma coleção de dados logicamente coerente com alguns significados inerentes aos mesmos.
- É projetado, construído e povoado com dados para um propósito específico.

Um banco de dados pode ser de qualquer tamanho e de complexidade variada. Pode ser gerado e mantido manualmente ou pode ser computadorizado. Um Sistema de Gerenciamento de Banco de Dados é uma coleção de programas que permitem que usuários criem e mantenham um banco de dados.

Um SGBD é também um sistema de software de propósito geral que facilita o processo de definição, construção e manipulação de banco de dados para várias aplicações. A definição de um banco de dados envolve especificação de tipos de dados, estruturas e restrições para que os dados possam ser armazenados no banco de dados.

A construção do banco de dados é o processo de armazenamento de dados em algum armazenamento médio que é controlado pelo SGBD. A manipulação do banco de dados inclui tais funções como consulta a banco de dados para recuperar dados específicos e a atualização destes dados. Chamaremos de Sistema de Banco de Dados, o banco de dados e o SGBD, juntos. A Figura 1 mostra a estrutura de um Sistema de Banco de Dados.



Abstração de Dados

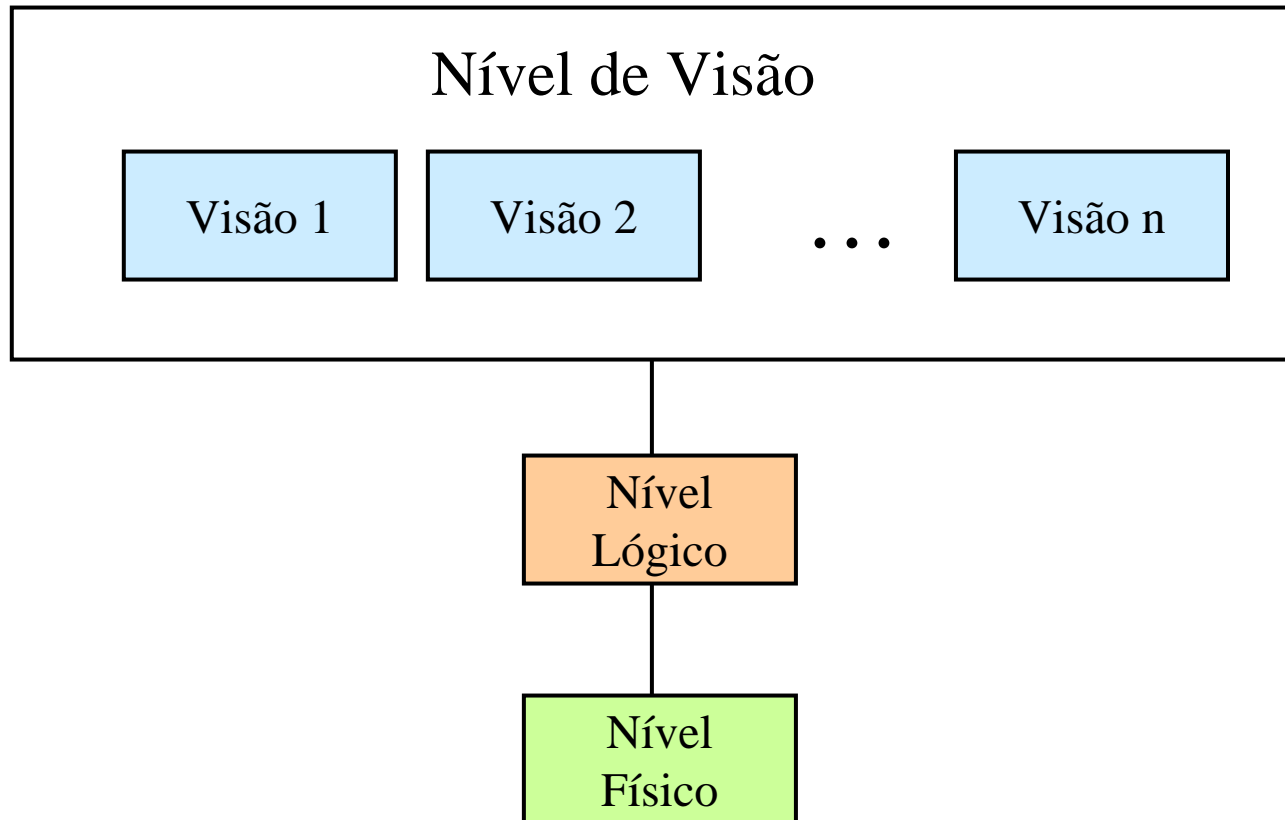
Um SGBD, como já dissemos, é uma coleção arquivos e programas inter-relacionados que permitem ao usuário o acesso para consultas e alterações a esses dados. Um dos grandes benefícios de um sistema de banco de dados é permitir aos usuários, uma **visão abstrata** dos dados.

Um sistema precisa ser eficiente na recuperação de informações. Esta eficiência muitas vezes está relacionada com a complexidade internas das estruturas de representação destes dados. Esta complexidade precisa ser transparente para os usuários, ou seja, através dos níveis de abstração, os SGBDs omitem a complexidade de armazenamento e manutenção dos dados que estão sendo manipulados, de modo a facilitar a interação dos usuários com o sistema.

- **Nível Físico:** é o nível de abstração mais baixo e descreve *como* os dados estão de fato armazenados.
- **Nível Lógico:** este é o nível de abstração médio, que descreve *quais* dados estão armazenados no banco de dados e quais são os inter-relacionamentos entre eles. O nível lógico é normalmente usado pelos administradores que precisam decidir quais as informações que precisam estar no banco de dados.

•**Nível de Visão:** Este nível de abstração mais alto que descreve apenas parte do banco de dados. Muitos usuários dos bancos de dados nem precisam conhecer todas as informações armazenadas. Para isso, níveis de visão são definidos de modo que as interações sejam simplificadas.

A Figura abaixo mostra o inter-relacionamento entre estes três níveis de abstração:



Podemos fazer uma analogia dos níveis de abstração com tipos de dados em linguagens de programação. Por exemplo, dada a estrutura abaixo:

```
Type cliente = record
```

```
    Nome_cliente : string;
```

```
    Seguro_social : string;
```

```
    Rua_cliente : string;
```

```
    Cidade_cliente : string;
```

```
end;
```

No nível físico, um registro de cliente pode ser escrito como um bloco consecutivo de memória. No nível lógico, cada registro é descrito por um tipo definido, como ilustrado no segmento de código acima. Normalmente os programadores e administradores do banco de dados trabalham neste nível de abstração. No nível de visão, os usuários vêem apenas um conjunto de programas de aplicação que escondem os detalhes dos tipos de dados.

Instâncias e Esquemas

Um banco de dados muda ao longo do tempo por meio das informações que nele são inseridas ou excluídas. O conjunto de informações contidas em determinado banco de dados em um dado momento é chamado **instância** do banco de dados. O projeto geral do banco de dados é chamado de **esquema**.

Fazendo uma análise comparativa com o exemplo dado de clientes, acima, uma variável poderia ser declarada como: **Var** cliente1 : cliente;

A definição de tipo em uma linguagem de programação corresponde ao **esquema** do banco de dados. O valor em um dado instante de uma variável de um determinado tipo, corresponde a uma **instância** do esquema do banco de dados.

Os sistemas de banco de dados apresentam diversos esquemas, referentes aos níveis de abstração que foram discutidos. Em geral os sistemas de banco de dados são suporte a um esquema físico, um esquema lógico e vários subesquemas.

A capacidade de modificar a definição dos esquemas em determinado nível, sem afetar o esquema do nível superior, é chamado independência de dados. Existem dois níveis de independência de dados:

- a) **Independência física de dados**
- b) **Independência lógica de dados**

Modelos de Dados

Um modelo de dados é um conjunto de ferramentas conceituais utilizadas para a descrição de dados, relacionamento entre esses dados, semântica de dados e regras de consistência. Os modelos são classificados em três diferentes grupos: **modelos lógicos com base em objetos**, **modelos lógicos com base em registros** e **modelos físicos**.

Modelos Lógicos com base em Objetos

Os modelos lógicos com base em objetos são usados na descrição de dados no nível lógico e de visões. Existem vários modelos desta categoria, tais como:

- Modelo Entidade-Relacionamento
- Modelo Orientado a Objetos
- Modelo Semântico de Dados
- Modelo Funcional de Dados

- **Modelo Entidade-Relacionamento**

O modelo de dados entidade-relacionamento (E-R) tem como base a percepção do mundo real como um conjunto de objetos básicos, chamados de entidades, e do relacionamento entre eles. Uma entidade é uma coisa ou um objeto do mundo real, que pode ser identificado por outros objetos. As entidades são descritas no banco de dados por meio de seus atributos. Um relacionamento é uma associação entre entidades. Além das entidades e dos relacionamentos, o modelo E-R representa certas regras as quais o conteúdo do banco de dados precisa respeitar.

- **Modelo Orientado a Objetos**

O modelo orientado a objetos, assim como o E-R, tem por base um conjunto de objetos. Um objeto contém valores armazenados em variáveis instâncias dentro do objeto. Um objeto também contém conjuntos de códigos que operam este objeto. Estes conjuntos de códigos são chamados de métodos. Os objetos que contém os mesmos tipos de valores e os mesmos métodos são agrupados em classes. Onde uma classe pode ser vista como uma definição de tipo para objetos.

Modelos Lógicos Baseados em Registros

Modelos lógicos baseados em registros são usados para descrever os dados no nível lógico e de visão. É usado tanto para especificar a estrutura lógica do banco de dados quanto para implementar uma descrição de alto nível. Dos três modelos apresentados, o modelo relacional é o que mais tem se destacado nos últimos anos. O modelo hierárquico e de rede é ainda usado em um grande número de banco de dados antigos.

- **Modelo Relacional**

O modelo relacional usa um conjunto de tabelas para representar tanto os dados como a relação entre eles. Cada tabela possui múltiplas colunas e cada uma possui um único nome.

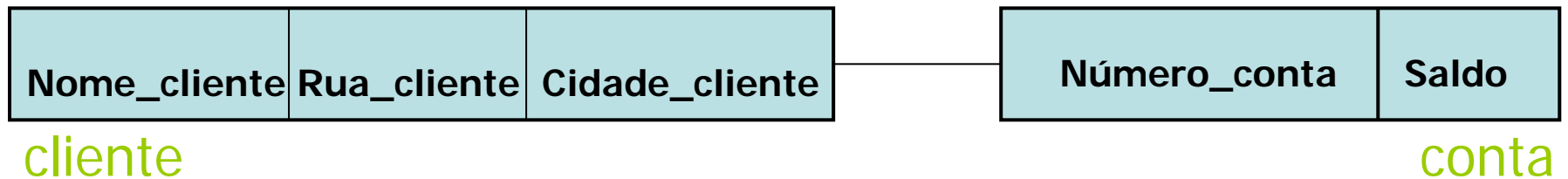
- **Modelo de Rede**

O modelo de rede representa os dados por um conjunto de registros e as relações entre esses registros são representadas por *links*, as quais podem ser vistas pelos ponteiros. Os registros são organizados no banco de dados por um conjunto arbitrário de gráficos.

Modelo de Rede

O modelo de rede é fortemente dependente da implementação. Muitas vezes é necessário criar registros artificiais para implementar relacionamentos. (Registros *dummy*)

O modelo de rede aumenta significativamente o trabalho do programador, tanto para o projeto de um banco de dados quanto para a manipulação de dados, se comparado ao modelo relacional.



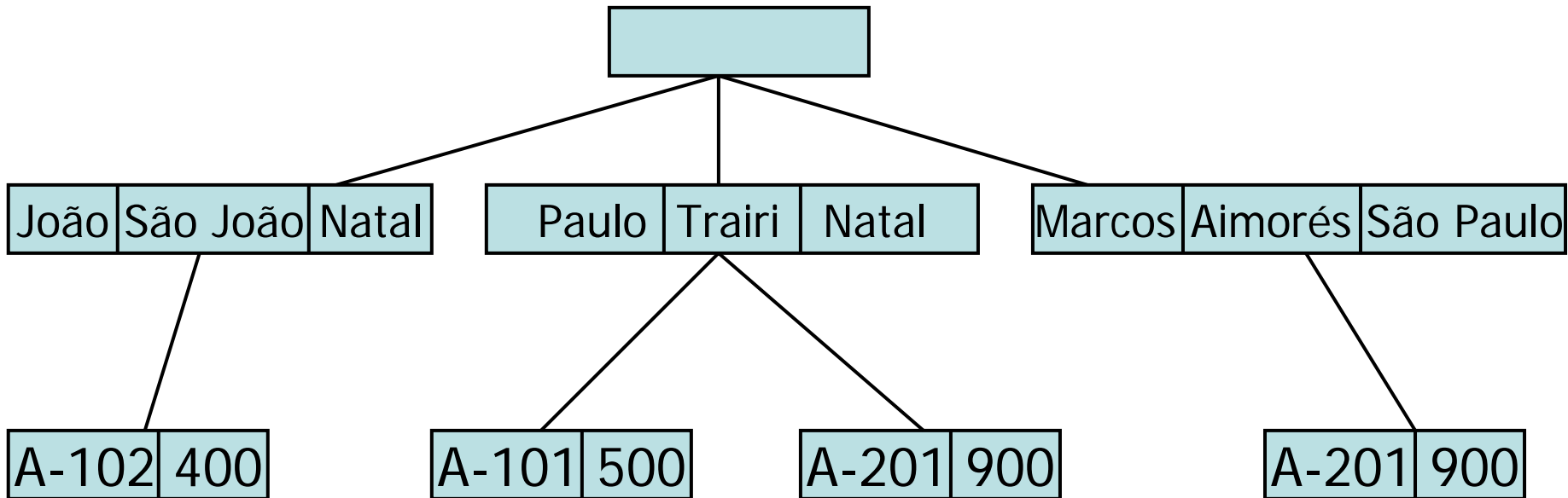
4.2.3 Modelo Hierárquico

O modelo hierárquico é similar ao modelo de rede pois os dados e suas relações são representados também por registros e links. A diferença é que no modelo hierárquico os registros estão organizados em árvores ao invés de gráficos arbitrários.

Todos os registros no modelo Hierárquico são organizados na forma de uma árvore com raiz, em que a raiz é um nó *dummy*.

O modelo hierárquico sofre dos mesmos problemas do modelo de rede. Por muito tempo, no entanto, os modelos hierárquico e de rede se mantiveram à frente do modelo relacional. Atualmente estes modelos vêm perdendo a importância.

Modelo Hierárquico



4.3 Modelos Físicos de Dados

Os modelos físicos de dados são usados para descrevê-los no nível mais baixo. Há poucos modelos físicos de dados em uso. Dois deles são conhecidos: o modelo unificado, e o modelo de partição de memória.