

Apresentação

Este curso tem como objetivo, oferecer uma noção geral sobre a construção de sistemas de banco de dados. Para isto, é necessário estudar modelos para a construção de projetos lógicos de bancos de dados, modelos para a construção de projetos físicos de banco de dados, técnicas de controle de dependência de dados e métodos de consultas.

Para construção dos modelos lógicos, será estudado o modelo Entidade Relacionamento, utilizando a abordagem proposta em [ELMAS89] que oferece uma notação rica em recursos, permitindo a modelagem de entidades normais, fracas, atributos simples, compostos, multivalorados, derivados e a modelagens de objetos mais complexos como classes e subclasses (modelo Entidade Relacionamento Extendido).

Para construção dos modelos físicos, será estudado o modelo Relacional como originalmente proposto por Codd.

Para eliminar dependência de dados, utilizaremos a normalização, abordando a 1^a, a 2^a, a 3^a Formas Normais, propostas originalmente por Codd.

Para a elaboração de consultas, será estudado a Álgebra Relacional, que nada mais é do que uma forma canônica para as linguagens de consulta e a linguagem de consultas SQL.

1. Introdução e Conceitos Gerais

A tecnologia aplicada aos métodos de armazenamento de informações vem crescendo e gerando um impacto cada vez maior no uso de computadores, em qualquer área em que os mesmos podem ser aplicados.

Um “banco de dados” pode ser definido como um conjunto de “dados” devidamente relacionados. Por “dados” podemos compreender como “fatos conhecidos” que podem ser armazenados e que possuem um significado implícito. Porém, o significado do termo “banco de dados” é mais restrito que simplesmente a definição dada acima. Um banco de dados possui as seguintes propriedades:

- um banco de dados é uma coleção lógica coerente de dados com um significado inerente; uma disposição desordenada dos dados não pode ser referenciada como um banco de dados;
- um banco de dados é projetado, construído e populado com dados para um propósito específico; um banco de dados possui um conjunto pré definido de usuários e aplicações;
- um banco de dados representa algum aspecto do mundo real, o qual é chamado de “mini-mundo” ; qualquer alteração efetuada no mini-mundo é automaticamente refletida no banco de dados.

Um banco de dados pode ser criado e mantido por um conjunto de aplicações desenvolvidas especialmente para esta tarefa ou por um “Sistema Gerenciador de Banco de Dados” (SGBD). Um SGBD permite aos usuários criarem e manipularem bancos de dados de propósito geral. O conjunto formado por um banco de dados mais as aplicações que manipulam o mesmo é chamado de “Sistema de Banco de Dados”.

1.1. Abordagem Banco de Dados X Abordagem Processamento Tradicional de Arquivos

1.1.1. Auto Informação

Uma característica importante da abordagem Banco de Dados é que o SGBD mantém não somente os dados mas também a forma como os mesmos são armazenados, contendo uma descrição completa do banco de dados. Estas informações são armazenadas no catálogo do SGBD, o qual contém informações como a estrutura de cada arquivo, o tipo e o formato de armazenamento de cada tipo de dado, restrições, etc. A informação armazenada no catálogo é chamada de “Meta Dados”. No processamento tradicional de arquivos, o programa que irá manipular os dados deve conter este tipo de informação, ficando limitado a manipular as informações que o mesmo conhece. Utilizando a abordagem banco de dados, a aplicação pode manipular diversas bases de dados diferentes.

1.1.2. Separação entre Programas e Dados

No processamento tradicional de arquivos, a estrutura dos dados está incorporada ao programa de acesso. Desta forma, qualquer alteração na estrutura de arquivos implica na

Sistemas de Bancos de Dados

alteração no código fonte de todos os programas. Já na abordagem banco de dados, a estrutura é alterada apenas no catálogo, não alterando os programas.

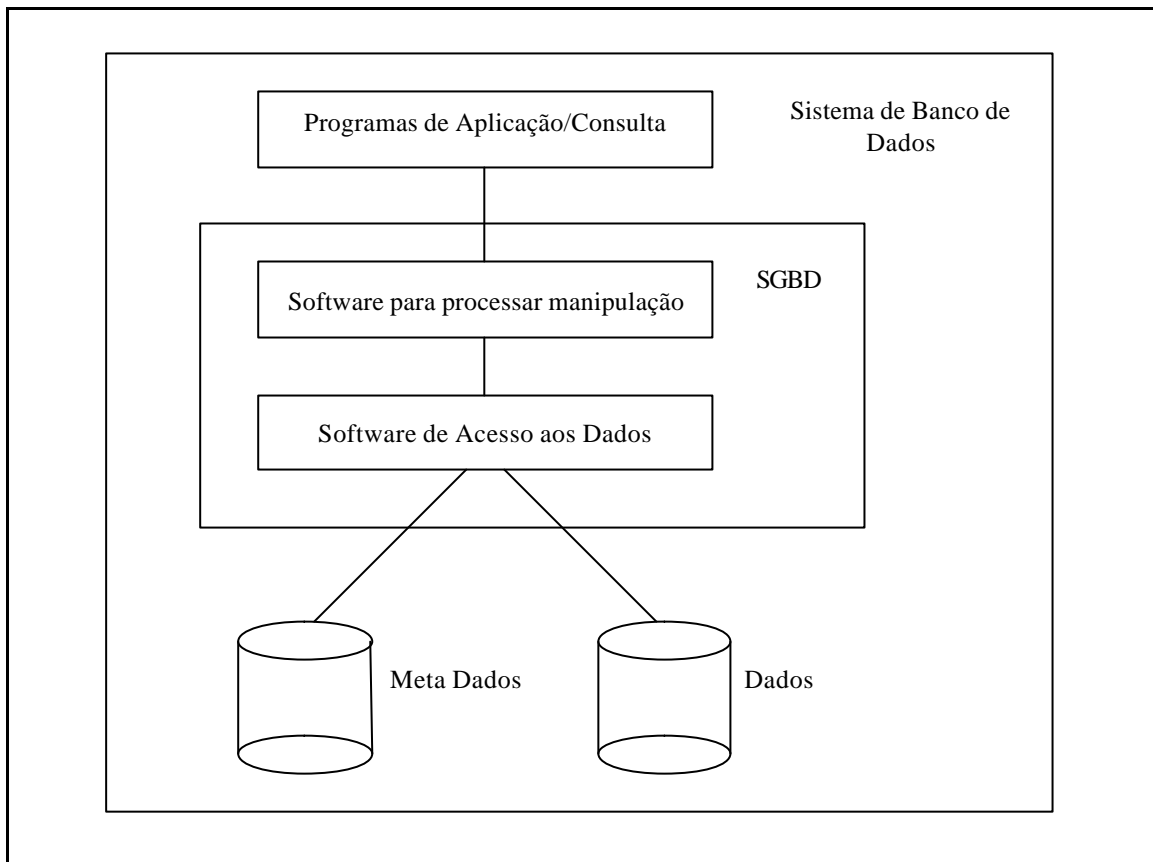


Figura 1. Um ambiente de Sistema de Banco de Dados

1.1.3. Abstração de Dados

O SGBD deve fornecer ao usuário uma “representação conceitual” dos dados, sem fornecer muitos detalhes de como as informações são armazenadas. Um “modelo de dados” é uma abstração de dados que é utilizada para fornecer esta representação conceitual utilizando conceitos lógicos como objetos, suas propriedades e seus relacionamentos. A estrutura detalhada e a organização de cada arquivo são descritas no catálogo.

1.1.4. Múltiplas Visões de Dados

Como um conjunto de informações pode ser utilizada por um conjunto diferenciado de usuários, é importante que estes usuários possam ter “visões” diferentes da base de dados. Uma “visão” é definida como um subconjunto de uma base de dados, formando deste modo, um conjunto “virtual” de informações.

1.2. Usuários

Para um grande banco de dados, existe um grande número de pessoas envolvidas, desde o projeto, uso até manutenção.

1.2.1. Administrador de Banco de Dados (DBA)

Sistemas de Bancos de Dados

Em um ambiente de banco de dados, o recurso primário é o banco de dados por si só e o recurso secundário o SGBD e os softwares relacionados. A administração destes recursos cabe ao Administrador de Banco de Dados, o qual é responsável pela autorização de acesso ao banco de dados e pela coordenação e monitoração de seu uso.

1.2.2. Projetista de Banco de Dados

O Projetista de Banco de Dados é responsável pela identificação dos dados que devem ser armazenados no banco de dados, escolhendo a estrutura correta para representar e armazenar dados. Muitas vezes, os projetistas de banco de dados atuam como “staff” do DBA, assumindo outras responsabilidades após a construção do banco de dados. É função do projetista também avaliar as necessidades de cada grupo de usuários para definir as visões que serão necessárias, integrando-as, fazendo com que o banco de dados seja capaz de atender a todas as necessidades dos usuários.

1.2.3. Usuários Finais

Existem basicamente três categorias de usuários finais que são os usuários finais do banco de dados, fazendo consultas, atualizações e gerando documentos:

- usuários casuais: acessam o banco de dados casualmente, mas que podem necessitar de diferentes informações a cada acesso; utilizam sofisticadas linguagens de consulta para especificar suas necessidades;
- usuários novatos ou paramétricos: utilizam porções pré-definidas do banco de dados, utilizando consultas preestabelecidas que já foram exaustivamente testadas;
- usuários sofisticados: são usuários que estão familiarizados com o SGBD e realizam consultas complexas.

1.2.4. Analistas de Sistemas e Programadores de Aplicações

Os analistas determinam os requisitos dos usuários finais e desenvolvem especificações para transações que atendam estes requisitos, e os programadores implementam estas especificações como programas, testando, depurando, documentando e dando manutenção no mesmo. É importante que, tanto analistas quanto programadores, estejam a par dos recursos oferecidos pelo SGBD.

1.3. Vantagens e desvantagens do uso de um SGBD

1.3.1. Controle de Redundância

No processamento tradicional de arquivos, cada grupo de usuários deve manter seu próprio conjunto de arquivos e dados. Desta forma, acaba ocorrendo redundâncias que prejudicam o sistema com problemas como:

- toda vez que for necessário atualizar um arquivo de um grupo, então todos os grupos devem ser atualizados para manter a integridade dos dados no ambiente como um todo;

- a redundância desnecessária de dados levam ao armazenamento excessivo de informações, ocupando espaço que poderia estar sendo utilizado com outras informações.

1.3.2. Compartilhamento de Dados

Um SGBD multi-usuário deve permitir que múltiplos usuários acessem o banco de dados ao mesmo tempo. Este fator é essencial para que múltiplas aplicações integradas possam acessar o banco.

O SGBD multi-usuário deve manter o controle de concorrência para assegurar que o resultado de atualizações sejam corretos. Um banco de dados multi-usuários deve fornecer recursos para a construção de múltiplas visões.

1.3.3. Restrição a Acesso não Autorizado

Um SGBD deve fornecer um subsistema de autorização e segurança, o qual é utilizado pelo DBA para criar “contas” e especificar as restrições destas contas; o controle de restrições se aplica tanto ao acesso aos dados quanto ao uso de softwares inerentes ao SGBD.

1.3.4. Representação de Relacionamentos Complexos entre Dados

Um banco de dados pode incluir uma variedade de dados que estão interrelacionados de várias formas. Um SGBD deve fornecer recursos para se representar uma grande variedade de relacionamentos entre os dados, bem como, recuperar e atualizar os dados de maneira prática e eficiente.

1.3.5. Tolerância a Falhas

Um SGBD deve fornecer recursos para recuperação de falhas tanto de software quanto de hardware.

1.3.6. Quando não Utilizar um SGBD

Em algumas situações, o uso de um SGBD pode representar uma carga desnecessária aos custos quando comparado à abordagem processamento tradicional de arquivos como por exemplo:

- alto investimento inicial na compra de software e hardware adicionais;
- generalidade que um SGBD fornece na definição e processamento de dados;
- sobrecarga na provisão de controle de segurança, controle de concorrência, recuperação e integração de funções.

Problemas adicionais podem surgir caso os projetistas de banco de dados ou os administradores de banco de dados não elaborem os projetos corretamente ou se as aplicações não são implementadas de forma apropriada. Se o DBA não administrar o banco de dados de forma apropriada, tanto a segurança quanto a integridade dos sistemas podem ser comprometidas. A sobrecarga causada pelo uso de um SGBD e a má administração justificam a utilização da abordagem processamento tradicional de arquivos em casos como:

Sistemas de Bancos de Dados

- o banco de dados e as aplicações são simples, bem definidas e não se espera mudanças no projeto;
- a necessidade de processamento em tempo real de certas aplicações, que são terrivelmente prejudicadas pela sobrecarga causada pelo uso de um SGBD;
- não haverá múltiplo acesso ao banco de dados.

2. Conceitos e Arquiteturas de um SGBD

2.1. Modelos de Dados

Uma das principais características da abordagem banco de dados, é que a mesma fornece alguns níveis de abstração de dados omitindo ao usuário final, detalhes de como estes dados são armazenados. Um “modelo de dados” é um conjunto de conceitos que podem ser utilizados para descrever a estrutura “lógica” e “física” de um banco de dados. Por “estrutura” podemos compreender o tipo dos dados, os relacionamentos e as restrições que podem recair sobre os dados.

Os modelos de dados podem ser basicamente de dois tipos:

- alto nível: ou modelo de dados conceitual, que fornece uma visão mais próxima do modo como os usuários visualizam os dados realmente;
- baixo nível: ou modelo de dados físico, que fornece uma visão mais detalhada do modo como os dados estão realmente armazenados no computador.

2.2. Esquemas e Instâncias

Em qualquer modelo de dados utilizado, é importante distinguir a “descrição” do banco de dados do “banco de dados” por si próprio. A descrição de um banco de dados é chamada de “esquema de um banco de dados” e é especificada durante o projeto do banco de dados. Geralmente, poucas mudanças ocorrem no esquema do banco de dados.

Os dados armazenados em um banco de dados em um determinado instante do tempo formam um conjunto chamado de “instância do banco de dados”. A instância altera toda vez que uma alteração no banco de dados é feita.

O SGBD é responsável por garantir que toda instância do banco de dados satisfaça ao esquema do banco de dados, respeitando sua estrutura e suas restrições. O esquema de um banco de dados também pode ser chamado de “intensão” de um banco de dados e a instância de “extensão” de um banco de dados.

2.3. A Arquitetura Três Esquemas

A principal meta da arquitetura “três esquemas” (figura 2) é separar as aplicações do usuário do banco de dados físico. Os esquemas podem ser definidos como:

- nível interno: ou esquema interno, o qual descreve a estrutura de armazenamento físico do banco de dados; utiliza um modelo de dados e descreve detalhadamente os dados armazenados e os caminhos de acesso ao banco de dados;
- nível conceitual: ou esquema conceitual, o qual descreve a estrutura do banco de dados como um todo; é uma descrição global do banco de dados, que não fornece detalhes do modo como os dados estão fisicamente armazenados;

- nível externo: ou esquema de visão, o qual descreve as visões do banco de dados para um grupo de usuários; cada visão descreve quais porções do banco de dados um grupo de usuários terá acesso.

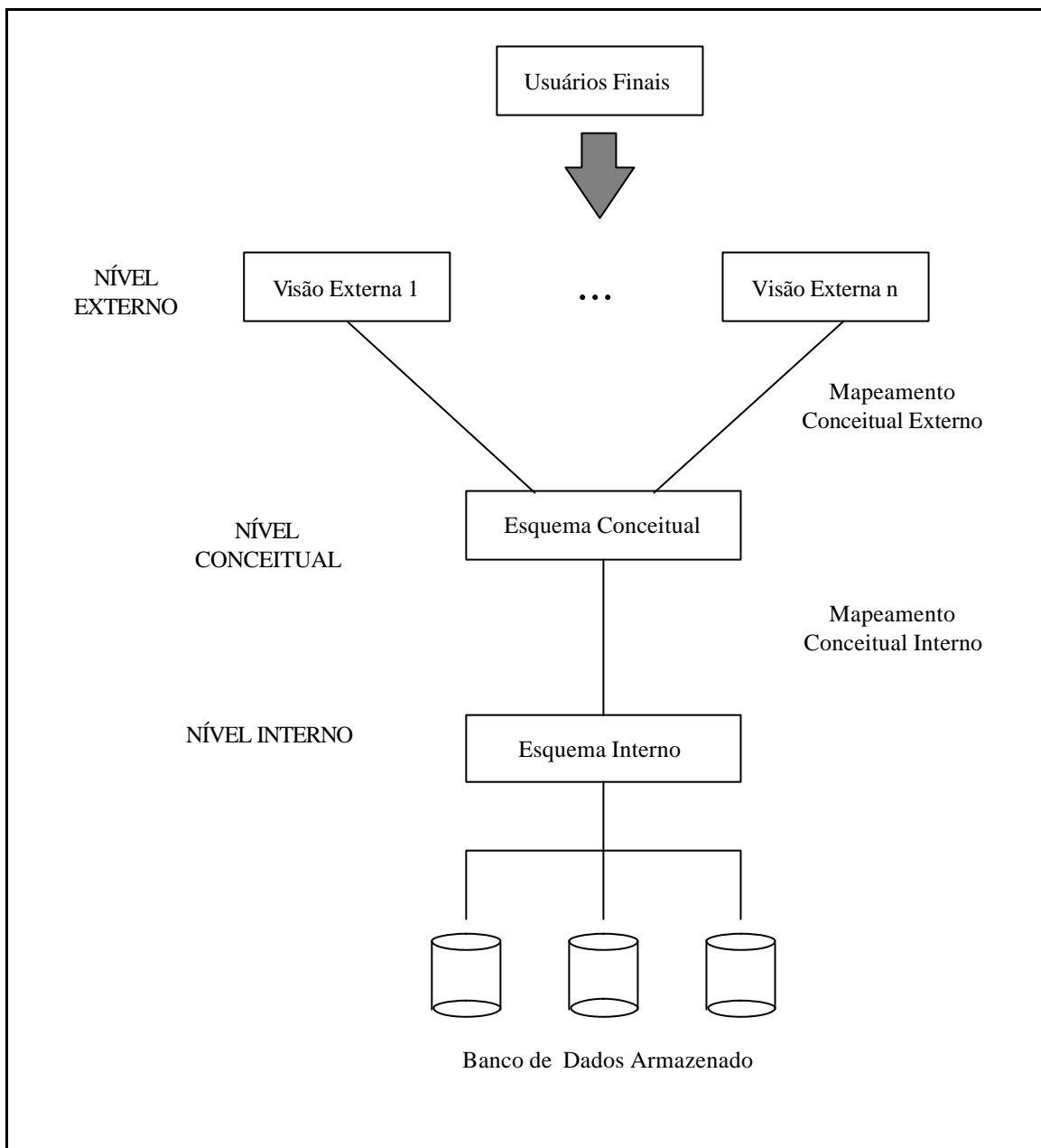


Figura 2: Arquitetura Três Esquemas

2.4. Independência de Dados

A “independência de dados” pode ser definida como a capacidade de se alterar um esquema em um nível em um banco de dados sem ter que alterar um nível superior (figura 2). Existem dois tipos de independência de dados:

- independência de dados lógica: é a capacidade de alterar o esquema conceitual sem ter que alterar o esquema externo ou as aplicações do usuário;

- independência de dados física: é a capacidade de alterar o esquema interno sem ter que alterar o esquema conceitual, o esquema externo ou as aplicações do usuário.

2.5. As Linguagens para Manipulação de Dados

Para a definição dos esquemas conceitual e interno pode-se utilizar uma linguagem chamada DDL (Data Definition Language - Linguagem de Definição de Dados). O SGBD possui um compilador DDL que permite a execução das declarações para identificar as descrições dos esquemas e para armazená-las no catálogo do SGBD. A DDL é utilizada em SGBDs onde a separação entre os níveis interno e conceitual não é muito clara.

Em um SGBD em que a separação entre os níveis conceitual e interno são bem claras, é utilizado uma outra linguagem, a SDL (Storage Definition Language - Linguagem de Definição de Armazenamento) para a especificação do esquema interno. A especificação do esquema conceitual fica por conta da DDL.

Em um SGBD que utiliza a arquitetura três esquemas, é necessária a utilização de mais uma linguagem para a definição de visões, a VDL (Vision Definition Language - Linguagem de Definição de Visões).

Uma vez que o esquema esteja compilado e o banco de dados esteja populado, usa-se uma linguagem para fazer a manipulação dos dados, a DML (Data Manipulation Language - Linguagem de Manipulação de Dados).

2.6. Os Módulos Componentes de um SGBD

Um SGBD é um sistema complexo, formado por um conjunto muito grande de módulos. A figura 3 mostra um esquema da estrutura de funcionamento de um SGBD.

2.7. Classificação dos SGBDs

O principal critério para se classificar um SGBD é o modelo de dados no qual é baseado. A grande maioria dos SGBDs contemporâneos são baseados no modelo relacional, alguns em modelos conceituais e alguns em modelos orientados a objetos. Outras classificações são:

- usuários: um SGBD pode ser mono-usuário, comumente utilizado em computadores pessoais ou multi-usuários, utilizado em estações de trabalho, mini-computadores e máquinas de grande porte;
- localização: um SGBD pode ser localizado ou distribuído; se ele for localizado, então todos os dados estarão em uma máquina (ou em um único disco) ou distribuído, onde os dados estarão distribuídos por diversas máquinas (ou diversos discos);
- ambiente: ambiente homogêneo é o ambiente composto por um único SGBD e um ambiente heterogêneo é o ambiente compostos por diferentes SGBDs.

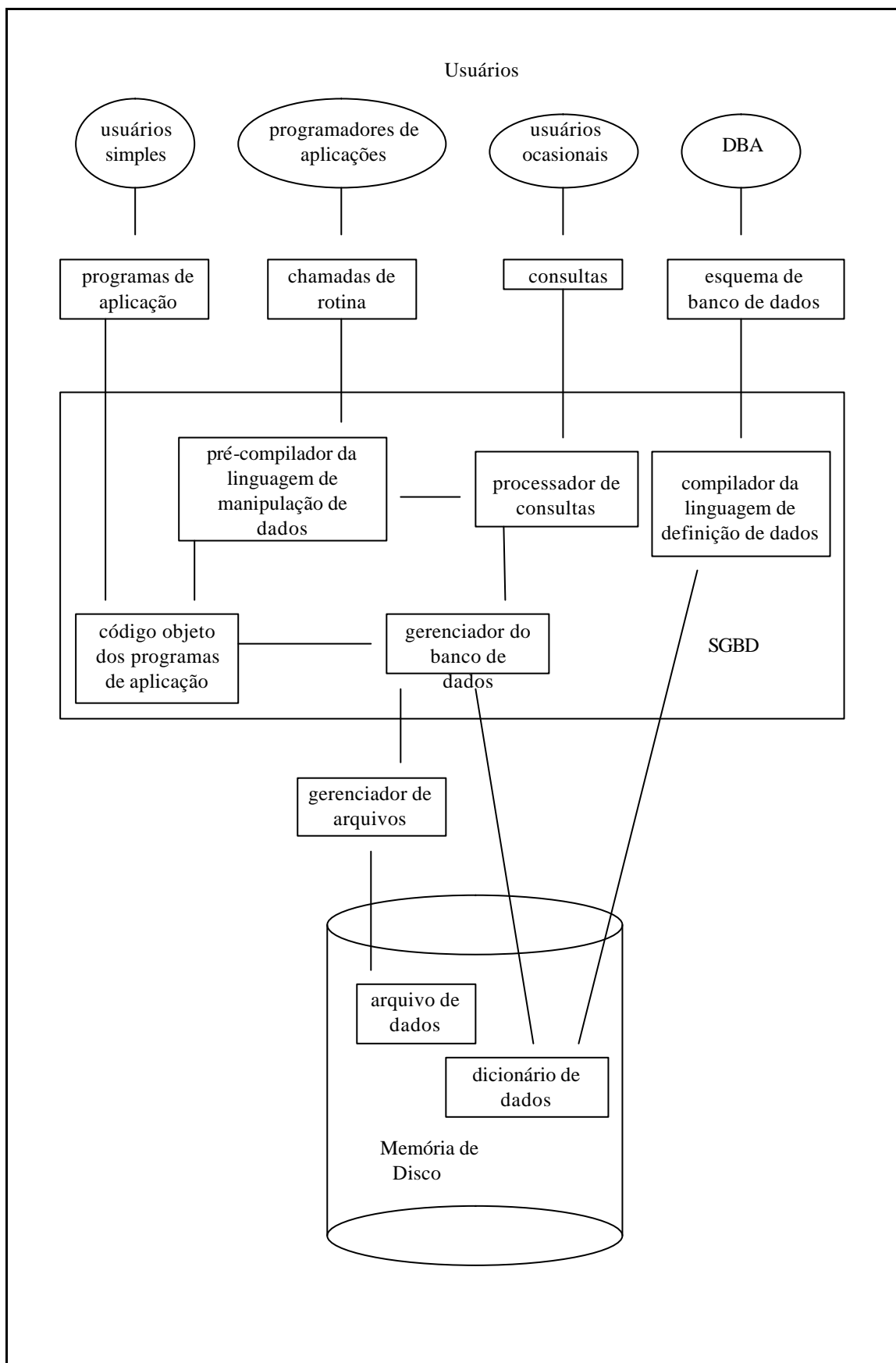


Figura 3: Estrutura de um Sistema Gerenciador de Banco de Dados