Armazenamento de Dados

Valéria Times

TÓPICOS

- Arquitetura de SGBD
- Hierarquia de Memória
- Discos
- Algoritmos Eficientes para Acesso ao Disco
- Otimização do Acesso ao Disco
- Falhas de Disco
- Totais de Verificação
- Armazenamento Estável
- Recuperação de Discos Danificados

Armazenamento de Dados

- **Aspecto Diferenciado:** Habilidade do SGBD para:
  - Lidar de forma eficiente com grandes volumes de dados.
  - Provê acesso transparente aos dados no disco.
    - Aplicações de BD não precisam se preocupar se os dados estão no disco ou na memória principal.

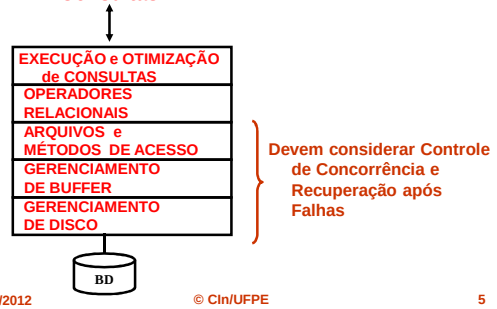
Armazenamento de Dados

- **Transferência de Dados:** SGBD mantêm dados em dispositivos de armazenamento secundário.
  - Quando um registro é solicitado para processamento, ele deve ser transferido do disco para memória principal:
    - Página contendo o registro é identificada pelo gerenciador de arquivos usando estruturas de dados auxiliares.
    - Gerenciador de arquivos solicita a página ao gerenciador de buffer
    - Gerenciador do buffer carrega a página solicitada ao gerenciador de disco na memória principal (buffer pool).

Armazenamento de Dados

- **Arquitetura de um SGBD:**

Consultas



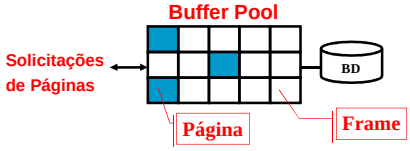
Armazenamento de Dados

- **Arquitetura de SGBD: Camadas Principais:**
  - **Gerenciamento de Disco:**
    - Gerencia o espaço de disco disponível.
    - Provê comandos para alocar (desalocar), gravar e ler uma página (unidade de dado).
    - Tamanho da página = tamanho do bloco.
    - Páginas são armazenadas como blocos de modo que a leitura/escrita de uma página possa ser feita em uma única operação de E/S.
    - Permite alocar uma seqüência contínua de blocos para dados que são freqüentemente acessados juntos.

Armazenamento de Dados

- Gerenciamento de Disco (Cont.):
  - Esconde detalhes de implementação e permite que outros componentes do SGBD visualize os dados como uma coleção de páginas.
  - Mantém informação sobre:
    - blocos currentemente em uso
    - quais blocos pertencem a quais páginas
    - blocos disponíveis
- Gerenciamento de Buffer:
  - Particiona a memória principal disponível em uma coleção de páginas (frames) localizadas no buffer pool.

Armazenamento de Dados



- Gerenciamento de Buffer (Cont.):
  - Objetiva trazer páginas do disco para memória principal quando necessário, em resposta às solicitações de leitura/gravação.
  - Outros componentes do SGBD não se preocupam se a página solicitada está ou não na memória.

Armazenamento de Dados

- Gerenciamento de Buffer (Cont.):
  - Outros componentes do SGBD:
    - Solicitam a página requerida
    - Informam ao gerenciador do buffer:
      - quando ela se torna desnecessária
      - se a página foi atualizada
  - Mantém duas variáveis principais para cada frame:
    - count: número atual de usuários da página.
    - dirty: indica se a página foi atualizada ou não desde sua transferência do disco.

Armazenamento de Dados

- Quando uma página é solicitada:
  - Verifica a sua existência no buffer pool. Em caso negativo:
    - Seleciona uma posição no buffer pool (frame) usando uma política de substituição de páginas.
    - Se a página a ser substituída tiver sido atualizada (dirty = TRUE):
      - Sua cópia no disco é atualizada.
      - Protocolo de recuperação pode requerer:
        - uso de registros de log (no buffer) para descrever as mudanças feitas à página.
        - Informação do log seja gravada ANTES da página ser gravada no disco.

Armazenamento de Dados

- Página solicitada é lida.
- Incrementa a variável count do frame contendo a página solicitada.

Armazenamento de Dados

- Gerenciamento de Buffer (Cont.):
  - Políticas de Substituição de Páginas:
    - Menos Recentemente Usada (LRU): registra a seqüência de acesso a todas as páginas.
    - Menos Freqüentemente Usada (LFU): registra o número de acessos a todas as páginas.
    - Ordem de Chegada (FIFO): registra a ordem de chegada das páginas.
  - A maioria dos SGBD usam uma variante da LRU, mas existem SGBD que adotam diferentes políticas em diferentes partições da memória.

Armazenamento de Dados

- Gerenciamento de Buffer (Cont.):
  - Não reusará um frame se sua variável count > 0
  - Se a página solicitada não estiver no buffer pool e nenhum frame estiver disponível:
    - um frame com count = 0 é escolhido para ser substituído
    - se todas tiverem count > 0, então a transação solicitante poderá ser abortada.

Armazenamento de Dados

- Gerenciamento de Buffer (Cont.):
  - Assume que a transação solicitante obteve bloqueio apropriado antes da página ser pedida
    - Controle de concorrência do SGBD garante que a transação tenha satisfeito sua solicitação de bloqueio antes dela solicitar uma página ao gerenciador de buffer para leitura/gravação.

Armazenamento de Dados

- Gerenciamento de Buffer em SGBD x SO:
  - Por que não usar a capacidade de memória virtual do SO:
    - SGBD pode prever mais precisamente a ordem na qual páginas serão acessadas (Page Reference Patterns).
      - Permite pré-carga de páginas
    - SGBD necessita de mais controle sobre quando uma página foi gravada no disco.
      - Garantir Consistência, Recuperação após Falhas, Controle de Concorrência.

Armazenamento de Dados

- Arquitetura de SGBD: Camadas Principais:
  - Arquivos e Métodos de Acesso:

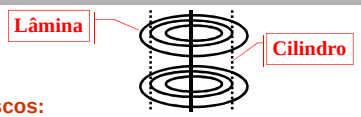
Coleção de páginas → Coleção de registros

  - Responsável por:
    - Manter as páginas de um arquivo.
    - Organizar os registros dentro das páginas.
    - Manter uma coleção de índices.

Armazenamento de Dados

- Hierarquia de Memória: Sistema de computador utiliza componentes de armazenamento que variam enormemente em:
  - Velocidade
  - Capacidade
  - Custo por bit
- Dispositivos:
  - primário (cache e memória principal)
  - secundário (disco)
  - terciário (fitas e CD-ROM).

Armazenamento de Dados

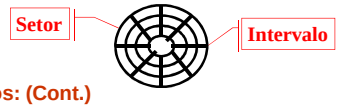


- Discos:
  - Suportam acesso direto à localização desejada e são bastante usados em aplicações de BD.
  - Têm várias lâminas circulares de material magnético, com trilhas concêntricas para armazenar bits.
  - Lâminas giram em torno de um eixo central.
  - Trilhas a uma certa distância radial do centro de uma lâmina formam um cilindro.

Armazenamento de Dados

- **Discos (cont.):**
  - Dados são mantidos em unidades lógicas de armazenamento (blocos) usadas por um aplicativo como SGBD.
  - Um array de cabeças de discos (uma por lâmina) move-se como uma unidade.
- **Blocos:**
  - Seqüência contínua de bytes usada para leitura/gravação de dados no disco.
  - São localizados em anéis concêntricos (trilhas), sobre uma ou mais lâminas.
  - Consistem em diversos setores consecutivos.

Armazenamento de Dados



- **Blocos: (Cont.)**
  - Para ler/escrever em um bloco, uma cabeça do disco deve estar posicionada em cima do mesmo.
  - Tamanho de um bloco corresponde a um valor múltiplo do tamanho do setor.
- **Setores:**
  - Trilhas são divididas em setores, que são separados por intervalos não magnetizados.

Armazenamento de Dados

- **Controlador de Disco:**
  - É um processador que controla uma ou mais unidades de disco.
  - Responsável pela movimentação das cabeças do disco até o cilindro apropriado para ler ou gravar a trilha solicitada.
  - Pode também programar solicitações concorrentes para acesso ao disco.
  - Coloca nos buffers os blocos a serem lidos ou gravados.

Armazenamento de Dados

- **Tempo de Acesso ao Disco:**
  - Tempo entre uma solicitação para leitura ou gravação de um bloco e o momento em que o acesso é concluído (latência do disco).
  - É influenciado por três fatores principais:
    - Tempo para mover as cabeças até o cilindro apropriado (tempo de busca).
    - Tempo para o disco girar até o primeiro dos setores que contêm o bloco desejado (latência rotacional).
    - Tempo enquanto o bloco se move sob a cabeça e é lido/gravado (tempo de transferência).

Armazenamento de Dados

- Dados devem estar na memória para SGBD processá-los.
- Unidade de transferência de dados entre o disco e a memória principal é um bloco.
  - Se um único item do bloco é necessário, o bloco inteiro é transferido.
  - A leitura/escrita de um bloco é chamada de operação de E/S.
- Tempo para ler/escrever um bloco varia, dependendo da localização do dado:

  Tempo Acesso = Tempo Busca + Latência Rotacional + Tempo Transferência

Armazenamento de Dados

- **Algoritmos de Acesso ao Disco:**
  - Quando os dados são volumosos e não cabem na memória principal, os algoritmos usados devem considerar que:
    - uma operação de E/S é em geral mais demorada do que o processamento dos dados em si.
  - Avaliação destes algoritmos se concentra portanto, no número de operações de E/S de disco necessárias.

Armazenamento de Dados

- Revisão do Algoritmo Merge-Sort:
  - Funciona pela mesclagem (ou intercalação) de listas classificadas para formar listas classificadas maiores.

Etapa	Lista1	Lista2	Saída
Início	1,3,4,9	2,5,7,8	nenhuma
1	3,4,9	2,5,7,8	1
2	3,4,9	5,7,8	1,2
3	4,9	5,7,8	1,2,3
4	9	5,7,8	1,2,3,4
5	9	7,8	1,2,3,4,5
6	9	8	1,2,3,4,5,7
7	9	nenhum	1,2,3,4,5,7,8
8	Nenhum	nenhum	1,2,3,4,5,7,8,9

Armazenamento de Dados

- Classificação por Intercalação de vários caminhos e duas fases:
  - Consiste em uma variante do Merge-Sort.
  - Método de classificação preferido na maioria das aplicações de BD.
  - Capaz de classificar enormes quantidades de dados no disco usando apenas duas leituras e duas gravações de disco para cada dado.
  - Realiza a mesclagem de várias sublistas ordenadas e consiste em duas fases:
    - Fase1: Classificação dos itens de dados
    - Fase2: Intercalação das sublistas

Armazenamento de Dados

- Classificação por Intercalação de vários caminhos e duas fases (Cont.):
  - Fase1:Classificação
    - Preenche a memória principal disponível com blocos da relação original a ser classificada.
    - Classifica os registros que estão na memória principal.
    - Grava os registros classificados provenientes da memória principal em novos blocos da memória secundária, formando uma única sublista classificada.

Armazenamento de Dados

- Classificação por Intercalação de vários caminhos e duas fases (Cont.):
  - Fase2:Intercalação
    - Mescla todas as sublistas classificadas em uma única lista ordenada.

Armazenamento de Dados

- Otimização do Acesso de Disco:
  - Pode-se tirar proveito da forma como os discos funcionam para:
    - tornar consultas mais rápidas
    - permitir ao sistema processar mais consultas ao mesmo tempo.
  - Tempo necessário para realizações de operações de BD é bastante afetado pela localização dos dados no disco.
    - Para minimizar este tempo, é importante posicionar os registros de forma estratégica no disco.

Armazenamento de Dados

- Exemplo: Se dois registros são freqüentemente acessados juntos, então eles devem ser posicionados juntos no disco.
  - mesmo bloco, mesma trilha, mesmo cilindro ou cilindro adjacente.
- Existem várias técnicas que visam melhorar o desempenho de um sistema de disco:
  - Organizar dados por cilindros
  - Utilizar vários discos em vez de um só
  - Espelhar discos
  - Usar um algoritmo de programação de discos
  - Realizar a busca prévia dos dados

Armazenamento de Dados

- **Organização baseada em Cilindros:**
  - Consiste em posicionar juntos no mesmo cilindro/trilha os blocos que serão acessados de forma que possa evitar:
    - tempo de busca
    - latência rotacional
  - **Vantagem:** Abordagem interessante para aplicativos em que os acessos podem ser previstos com antecedência, e apenas um processo estiver usando o disco.
  - **Desvantagem:** Nenhuma ajuda para aplicativos em que os acessos são imprevisíveis.

Armazenamento de Dados

- **Vários Discos:**
  - Consiste em dividir os dados entre vários discos pequenos em lugar de um único disco grande para permitir o acesso paralelo.
  - Existência de um número maior de conjuntos de cabeças capazes de percorrer os blocos independentemente pode aumentar o número de acessos a blocos por unidade de tempo.
  - **Vantagem:** Aumenta a taxa em que solicitações de leitura/gravação podem ser satisfeitas.

Armazenamento de Dados

- **Vários Discos (Cont.):**
  - **Desvantagens:**
    - Solicitações de leitura/gravação no mesmo disco não podem ser satisfeitas ao mesmo tempo (problema de colisão de acessos).
      - Fator aceleração pode ser menor que o fator pelo qual aumenta o número de discos.
    - Custo total de vários discos pequenos pode exceder o custo de um único disco com a mesma capacidade de armazenamento.

Armazenamento de Dados

- **Espelhamento:**
  - Consiste em criar duas ou mais cópias dos dados em discos isolados para permitir o acesso paralelo.
  - Permite acessar diversos blocos de uma só vez.
  - **Vantagens:**
    - Aumenta a taxa em que solicitações de leitura/gravação podem ser satisfeitas.
    - Melhora a tolerância a falhas porque protege os dados no caso de um dos discos apresentar problemas.

Armazenamento de Dados

- **Espelhamento (Cont.):**
  - **Vantagens:**
    - Não exibe o problema de colisão de acessos.
  - **Desvantagem:**
    - Necessidade de aquisição de dois ou mais discos, mas obtendo a capacidade de armazenamento de apenas um.

Armazenamento de Dados

- **Programação do Disco:**
  - Consiste em usar um algoritmo de programação de discos (no SO, SGBD ou no controlador de discos) para selecionar a ordem na qual diversos blocos solicitados serão lidos/gravados.
  - Um modo simples e eficiente de programar grandes números de solicitações de blocos é conhecido como algoritmo do elevador.

Armazenamento de Dados

- Programação do Disco (cont.):
  - Algoritmo do elevador: Otimiza os acessos:
    - Enfileirando as solicitações de acesso.
    - Tratando-as em uma ordem que permita às cabeças fazerem um única varredura através do disco.
    - Cabeças interrompem o tratamento de uma solicitação toda vez que alcançam um cilindro contendo um ou mais blocos com solicitações de acesso pendentes.

Armazenamento de Dados

- Programação do Disco (cont.):
  - Algoritmo do elevador
  - Vantagem: Reduz o tempo médio para leitura/gravação de blocos quando os acessos a blocos são imprevisíveis.
  - Desvantagem: É menos efetivo em situações nas quais existem poucas solicitações de acesso a disco esperando.

Armazenamento de Dados

- Busca Prévia (Bufferização Dupla):
  - Consiste na leitura/gravação de trilhas/cilindros inteiros em conjunto.
  - Realiza a transferência prévia dos blocos para a memória principal, prevendo seu uso posterior.
  - Vantagem: Acelera o acesso quando os blocos necessários são conhecidos, mas a sincronização das solicitações é dependente dos dados.
  - Desvantagens:
    - Exige buffers extras na memória principal.
    - Não oferece nenhuma ajuda quando os acessos são aleatórios.

Armazenamento de Dados

- Modos de Falha de Disco:
  - Para evitar perda de dados, os sistemas devem ser capazes de tratar erros.
  - Principais tipos de falhas de disco são:
    - Falha Intermitente: Consiste em um erro de leitura/gravação que não voltará a ocorrer se a operação for repetida (mais comum).
    - Permanente: Alguns dados do disco estão danificados e não podem mais ser lidos corretamente (decadência da mídia).

Armazenamento de Dados

- Modos de Falha de Disco (cont.):
  - Falha de Gravação: Tentativa mal sucedida de gravar um setor e de recuperar o setor anteriormente gravado.
  - Disco Danificado: Disco inteiro se torna ilegível, de modo repentino e permanente (queda de disco).

Armazenamento de Dados

- Totais de Verificação:
  - Consiste em uma técnica usada para determinar o *status* Bom/Ruim de um setor em uma operação de leitura.
  - Cada setor tem alguns bits adicionais (total de verificação) definidos de acordo com os valores dos bits de dados armazenados neste setor.
  - Baseia-se no uso de bits de paridade, onde:
    - Bit de Paridade: um bit extra para tornar par o número de valores 1 em uma seqüência de bits.

Armazenamento de Dados

- **Totais de Verificação (Cont.):**
  - Com a verificação de paridade, falhas intermitentes e permanentes podem ser detectadas, embora não possam ser corrigidas.

Armazenamento de Dados

Totais de Verificação

Existência de uma falha de mídia ou falha de leitura/gravação

Corrigir o erro

- Em operações de gravação, podemos nos encontrar em uma situação onde:
  - sobrescrevemos o conteúdo anterior de um setor e
  - ainda não podemos ler o novo conteúdo.

Armazenamento de Dados

- **Armazenamento Estável:** Consiste na:
  - Realização de duas cópias de todos os dados
  - Cuidado em relação à ordem em que estas cópias são gravadas.
  - **Vantagem:** Um único disco pode ser usado para proteger contra quase todas as falhas permanentes de um único setor.
  - **Desvantagem:** Custo de armazenamento é duplicado.

Armazenamento de Dados

- **Recuperação de Discos Danificados:**
  - Forma mais séria de falhas de disco, onde os dados são permanentemente destruídos.
  - Existem vários esquemas que visam reduzir o risco de perda de dados por discos danificados.
  - Estendem idéia de verificação de paridade ou de setores duplicados.
  - Expressão comum para esta classe de estratégias é **RAID** (**Redundant Array of Independent Disks**)
    - Array Redundante de Discos Independentes