COLÉGIO ESTADUAL PROTÁSIO ALVES
Av. Ipiranga, 1090 – Cep 90160-091 – Porto Alegre-RS.

Professora: Maria Helena – Maria-hsilva368@educar.rs.gov.br

AULA 3 - GERÊNCIA DE MEMÓRIA

Objetivo: Identificar as principais estratégicas utilizadas SO para otimizar o uso e gerenciamento da memória no sistema Multiprogramados . Conhecer a estrutura hierárquica e forma de acesso ao programas e processador.

Introdução

A memória é um importante recurso que deve ser gerenciado cuidadosamente. É senso comum perceber que, embora a capacidade de memória disponível nos sistemas de computação cada vez aumente mais, os desenvolvedores de software demandam mais memória para que seus programas possam ser armazenados e executados. A questão é que a memória, ao contrário do processador, possui um limite e por isso seu gerenciamento é crítico para o funcionamento dos processos pelo SO. Gerenciamento de memória é a tarefa desempenhada pela parte do SO que controla o uso da memória.

Funções da Gerência de Memória

É função dessa parte do SO é conhecer quais regiões da memória estão em uso e quais não estão sendo usadas, alocar memória para processos quando eles necessitarem e desalojá-los quando os processos terminarem de ser executados, gerenciar o swapping entre a memória principal e o disco, quando a memória principal não for grande o suficiente para comportar todos os processos.

A Gerência de Memória em Sistemas Multiprogramados

A necessidade de manter múltiplos programas ativos na memória do sistema impõe outra necessidade: a de controlar como esta memória é utilizada por estes vários programas. O gerenciamento de memória é, portanto, o resultado da aplicação de duas práticas distintas dentro de um sistema de computação:

• Como a memória é vista, isto é, como pode ser utilizada pelos processos existentes neste sistema.

• Como os processos são tratados pelo SO quanto às suas **necessidades de uso de memória.**

A Organização Hierárquica da Memória

Em um sistema de computação, o armazenamento de dados ocorre em diversos níveis.

Isso quer dizer que o armazenamento é realizado em diferentes tipos de dispositivos devido a quatro fatores básicos:

1. Tempo de acesso.
2. Velocidade de operação.
3. Custo por unidade de armazenamento.
4. Capacidade de armazenamento.

Com isso, o projetista de um sistema operacional determina quanto de cada tipo de memória será necessário para que o sistema seja, ao mesmo tempo, eficiente e economicamente viável. Ocorre que quanto mais rápida mais cara a memória e menor sua capacidade de armazenamento de dados. Abaixo é exibida uma figura demonstrando essa hierarquia da organização da memória:

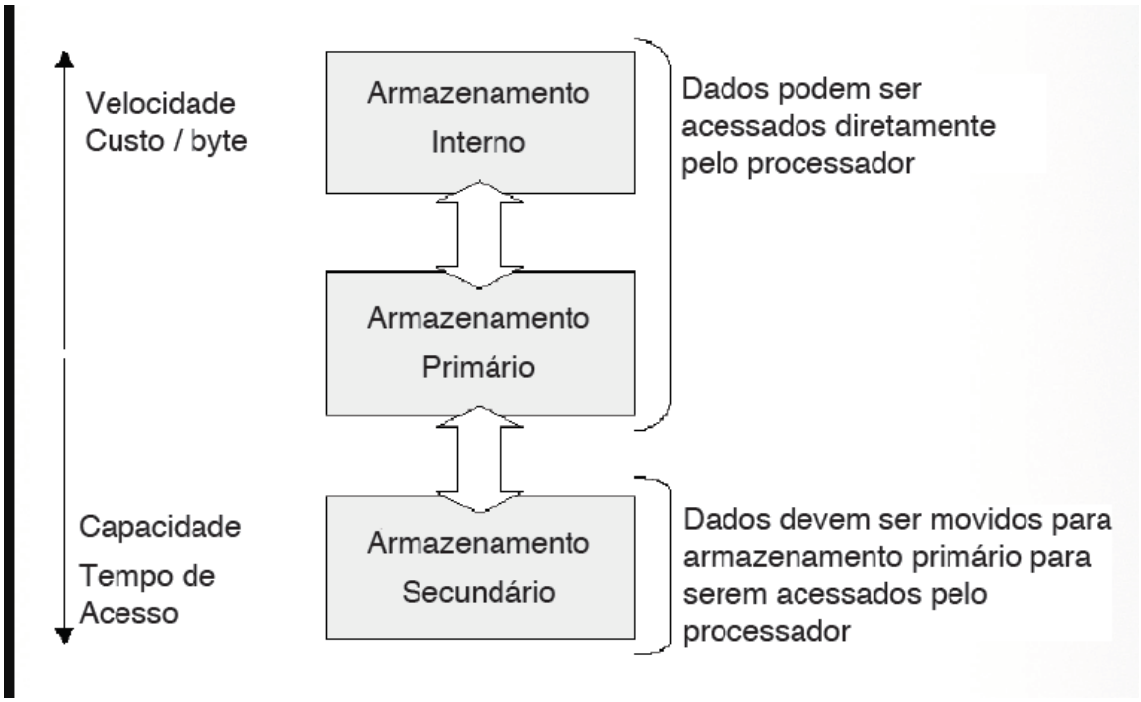

Figura 1- Gerenciamento - Organização da Memória em Níveis

Gerenciamento da memória em níveis

O Armazenamento Interno são posições de memória disponíveis internamente ao processador para permitir ou agilizar sua operação. É constituído dos registradores do processador e de seu cache interno.

O Armazenamento Primário são as posições de memória interna diretamente acessível ao processador. Tipicamente, são CIs de memória SRAM, DRAM, EPROM, PROM, entre outras.

O Armazenamento Secundário são as posições de memória externa que não podem ser acessadas diretamente pelo processador, devendo ser movidas para o Armazenamento Primário antes da sua utilização. Tipicamente, são os dispositivos de armazenamento de massa tal como o disco rígido.

Perceba que o Armazenamento Interno possui as maiores velocidades de acesso, ou seja, os menores tempos de acesso, representando os melhores dispositivos em termos de desempenho, embora sejam os mais caros. Por outro lado, os dispositivos de Armazenamento Secundário são os de maior capacidade e os de melhor relação custo/Byte, mas consideravelmente mais lentos.

O Armazenamento Primário representa um caso intermediário, em que a velocidade e o tempo de acesso são adequados à operação direta com o processador, mas cujo custo ainda assim é alto.

Com a evolução dos computadores, a atual organização conta com outros elementos adicionados para otimizar o desempenho do sistema e, ainda assim, reduzir seu custo.

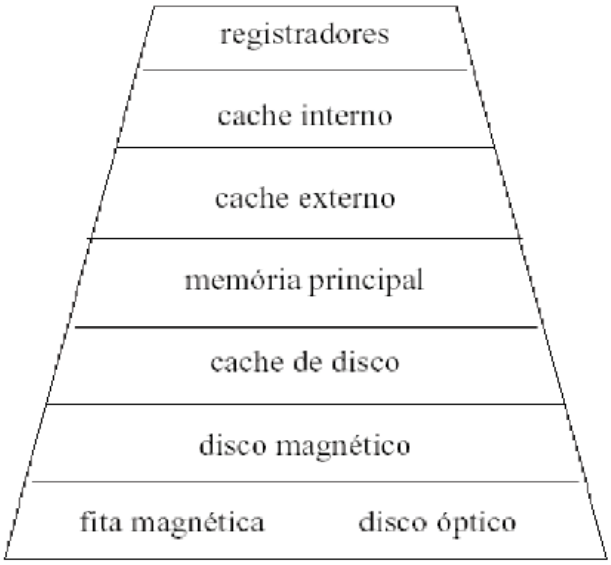

Figura 2- Gerenciamento de Memória - Hierarquia da Memória

Os registradores, implementados em número limitado em razão de seu custo, são geralmente usados para manter dentro do processador dados frequentemente utilizados.

Os caches interno e externo, em razão de sua maior velocidade, são usados para manter uma porção do programa que pode ser executados mais rapidamente do que na memória principal, aumentando o desempenho do sistema.

A memória primária armazena os programas e dados em execução no sistema.

Os dispositivos de armazenamento Secundário são usados para preservação dos dados de forma perene. O cache de disco é utilizado para acelerar a operação das unidades de disco, podendo esta técnica ser utilizada para outros tipos de periféricos.

Tipos de Gerenciamento de Memória

De maneira geral, sistemas de gerenciamento de memória podem ser divididos em duas classes:

- aqueles que movem processos (programas) do disco para a memória principal e vice-versa.
- aqueles que não realizam isto, trabalhando somente na memória.

Alocações Particionadas Estática e Dinâmica

Em sistemas Multiprogramados, a memória primária é dividida em blocos chamados de partições. Inicialmente, as partições, embora de tamanho fixo, não tinham necessariamente o mesmo tamanho entre elas, possibilitando diferentes configurações para sua utilização. Este esquema é conhecido como alocação particionada estática e tinha como grandes problemas:

O fato de os programas, normalmente, não preencherem totalmente as partições onde eram carregados, desperdiçando espaço.

Se um programa fosse maior do que qualquer partição livre, ele ficaria aguardando uma que o acomodasse, mesmo se existisse duas ou mais partições adjacentes que, somadas, totalizassem o tamanho do programa. Este tipo de problema, onde pedaços de memória ficam impedidos de serem usados por outros programas, é chamado de fragmentação.

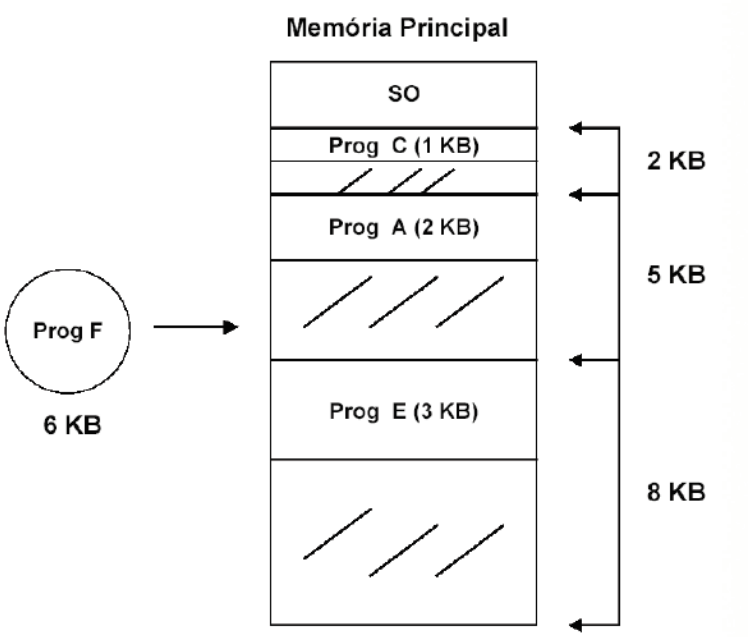

Figura 3- Gerenciamento de Memória - Problema de Fragmentação da Memória

Dado o problema da fragmentação na alocação particionada estática, foi necessário outro tipo de alocação como solução e, conseqüentemente, o aumento do compartilhamento da memória. Na alocação particionada dinâmica, foi eliminado o conceito de partições de tamanho fixo. Nesse esquema, cada programa utilizaria o espaço que necessitasse, passando esse bloco a ser sua partição.

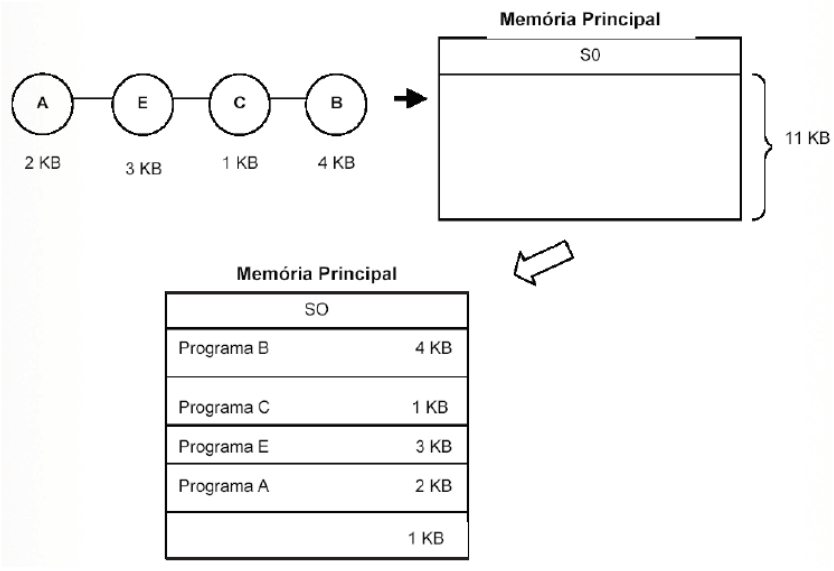

Figura 4- Gerenciamento de Memória - Esquema de alocação particionada dinâmica da Memória

A princípio, o problema da fragmentação pareceu estar resolvido, porém, neste caso, a fragmentação começará a ocorrer, realmente, quando os programas forem terminando e deixando espaços cada vez menores na memória, não permitindo o ingresso de novos programas. Para exemplificar este fato, veja os seguintes programas; com o término de B e E, mesmo existindo 8 KB livres de memória, o programa D, de 6 KB, não poderá ser carregado.

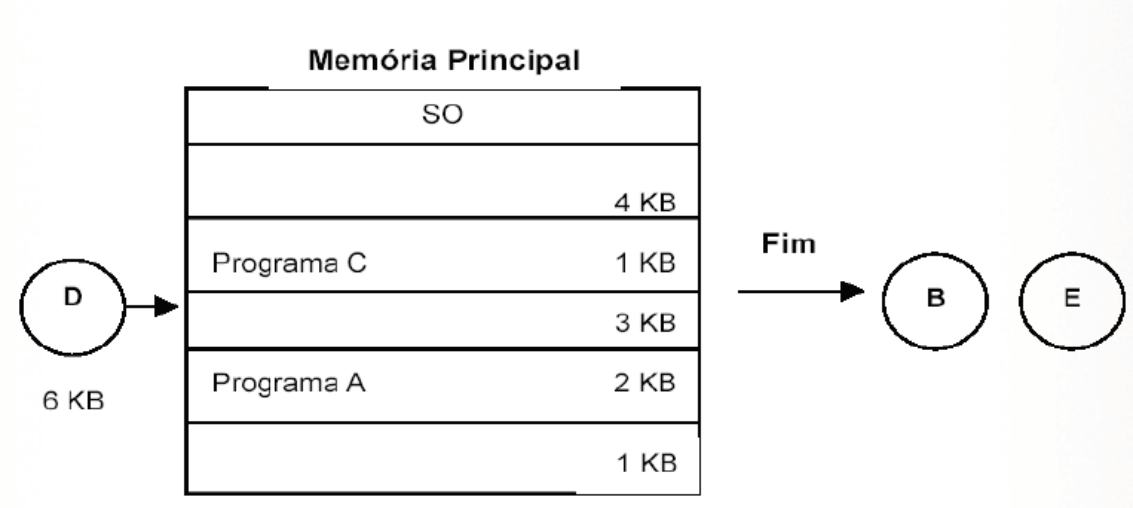

Figura 5- Gerenciamento de Memória - Memória Principal

Depois de já ter sido detectada a fragmentação da memória, há duas soluções para o problema:

a) Caso o programa C termine, o sistema pode reunir apenas os espaços adjacentes, produzindo espaço de tamanho 8 KB.

b) Caso o programa C continue executando, o sistema pode realocar todas as partições ocupadas, eliminando todos os espaços entre elas e criando uma única área livre contígua.

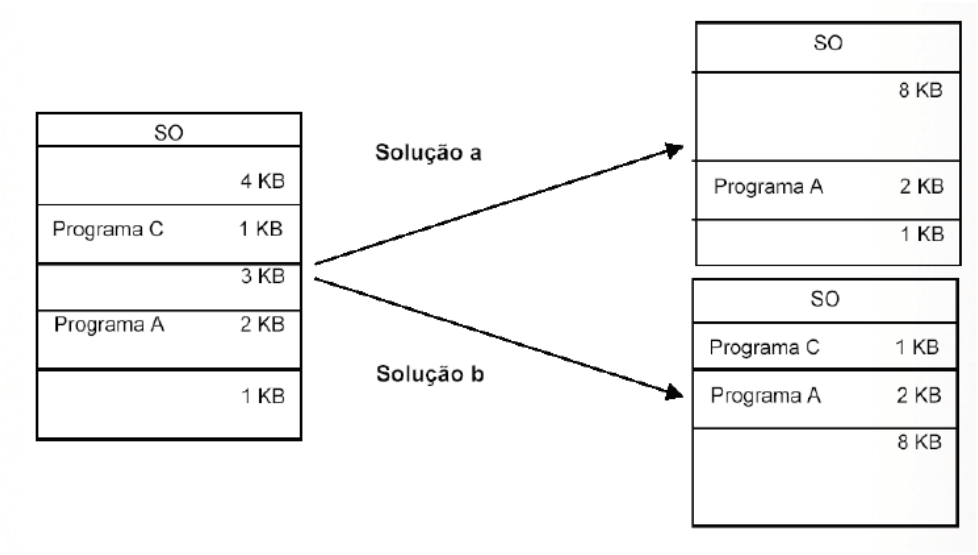

Figura 6- Gerenciamento de Memória - Soluções Implementadas

A complexidade do algoritmo de desfragmentação e o consumo de recursos do sistema, como processador e área em disco, podem tornar este processo inviável. É importante perceber que, nesses dois tipos de gerenciamento de memória apresentados, o espaço de endereçamento é igual ao tamanho da memória primária existente no sistema.

A técnica de Swapping

Em um sistema de Processamento em Lotes, a organização de memória em partições fixas é simples e eficiente. Desde que jobs suficientes possam ser mantidos na memória, de modo que a CPU fique ocupada todo o tempo, não existe razão para usar outra organização mais complexa.

Em sistemas de tempo compartilhado, a situação é diferente: normalmente existem mais usuários do que memória para manter todos os processos (programas), de modo que é necessário manter os processos em excesso no disco. Para executar tais processos, é necessário que eles sejam trazidos para a memória principal. O movimento de processos da memória principal para o disco e vice-versa é denominado swapping.

Nas alocações particionadas estática e dinâmica, um programa permanecia na memória principal até o final da sua execução, inclusive nos momentos em que esperava um evento, como uma operação de leitura ou gravação em periféricos. Em outras palavras, o programa somente sairia da memória principal quando tivesse terminada sua execução.

A técnica de swapping pode ser usada em sistemas Multiprogramados com partições de tamanho variável. Desta forma, de acordo com algum critério, um programa pode ser movido da memória principal para o disco (swap out) e este mesmo programa pode voltar do disco para a memória principal (swap in), como se nada tivesse acontecido.

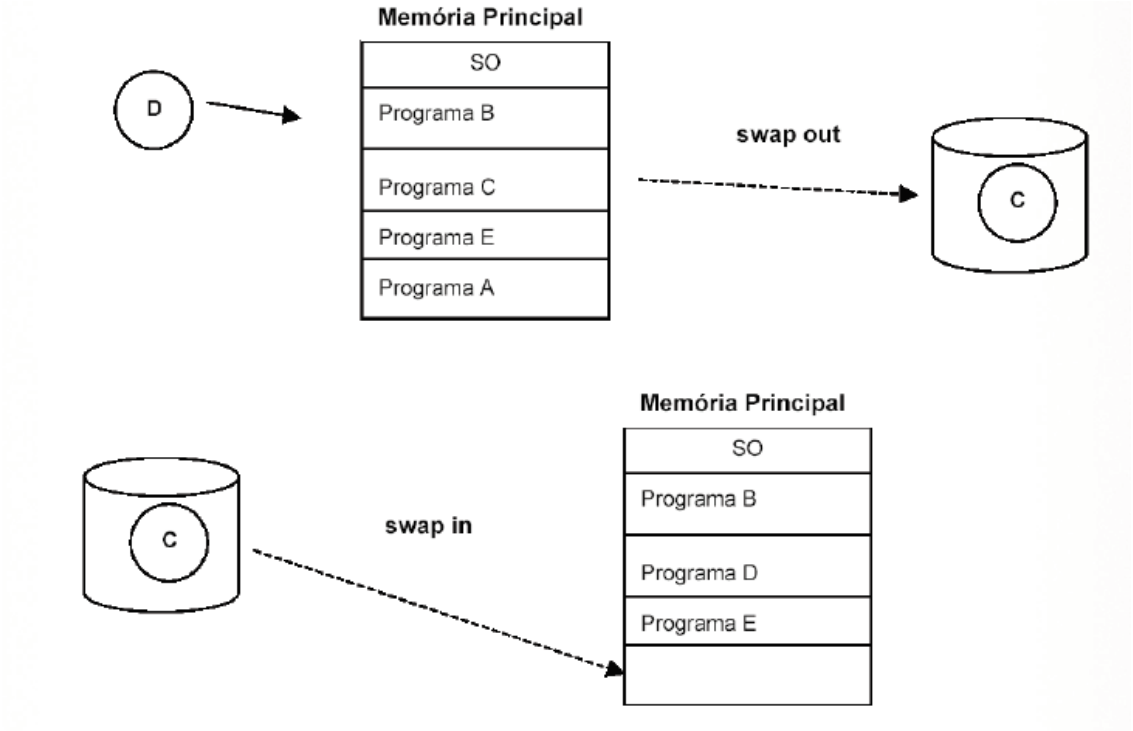

Figura 7- Gerenciamento de Memória - Técnica de Swapping

Para um maior aprofundamento no assunto, é relevante mencionar que o swapping é realizado por rotinas especiais do SO chamadas de relocadores ou swappers. A existência de relocadores em um sistema depende do tipo de gerenciamento de memória oferecido pelo SO. Uma explicação simplificada do trabalho realizado pelos relocadores é apresentada a seguir.

Seguindo instruções do SO, que detém o gerenciamento de memória e dos processos, um relocador pode ser comandado para retirar o conteúdo de uma área de memória, armazenando-a em disco. O que geralmente ocorre é que o relocador realiza uma cópia desta área de memória em um arquivo especial denominado arquivo de troca ou swap file. Ao copiar a área de memória para o disco, tal área é assinalada como livre, tornando disponível para outros processos. Também, é efetuado um registro do que foi copiado para a memória possibilitando a recuperação deste conteúdo.

Memória Virtual

O conceito de relocação de memória possibilitou o desenvolvimento de um meio mais otimizado de utilização de memória chamado memória virtual. O conceito de memória virtual está fundamentado em desvincular o endereçamento feito pelo programa dos endereços físicos da memória principal. Assim, os programas e suas estruturas de dados deixam de estar limitados ao tamanho da memória física disponível.

O termo memória virtual é normalmente associado com a habilidade de um sistema endereçar muito mais memória do que a fisicamente disponível. Este conceito surgiu em 1960 no computador Atlas, construído pela Universidade de Manchester (Inglaterra), embora sua utilização mais ampla tenha acontecido recentemente.