## Gerenciamento Da Memória

### 1. Introdução

Basicamente, o gerenciamento da memória é a ferramenta utilizada para permitir aos programas em execução no computador utilizarem a memória do computador para armazenar as instruções e dados que serão manipulados. Os principais objetivos a serem alcançados no gerenciamento da memória são:

* Oferecer uma área de armazenamento para os processos serem executados;
* Proteger a execução dos processos contra eventuais falhas causadas por erro em um determinado processo;
* Criar um ambiente de execução com desempenho satisfatório aos usuários;

Além dessas características, adicionalmente deseja-se de um bom gerenciador:

* Compartilhamento da memória entre os processos (em sistemas avançados);
* Permitir aos programadores desenvolverem programas que tenham um acesso transparente à memória do computador.

Nos sistemas operacionais modernos, o gerenciamento da memória realizará essas tarefas através da Proteção da Memória.

### 2. Proteção da Memória

A Proteção da Memória é a estratégia utilizada pelos sistemas operacionais para controlar a memória em utilização no computador. A principal função da proteção da memória é a separação dos processos e suas informações, e impedir que um processo acesse uma área de memória que não foi reservada para ele. Para realizar a proteção da memória, uma das seguintes medidas pode ser escolhida: Segmentação da Memória ou a Paginação da Memória.

Na Paginação da Memória, a memória é dividida em páginas, com tamanho fixo. Com a utilização de outro mecanismo chamado Memória Virtual, cada página pode ser armazenada em qualquer lugar da memória do computador. A Segmentação da Memória é a divisão da memória do computador em segmentos de dados. Quando um programa vai acessar uma área de memória, ele deve informar também o número do segmento de memória.

### 3. Compartilhamento da Memória

Uma memória é chamada de “memória compartilhada” quando vários processos podem acessá-la simultaneamente. Dependo do processador utilizado no computador, vários processos podem ser executados ao mesmo tempo, e a comunicação entre os processos pode ser feita utilizando-se uma área da memória que seja acessível a todos os processos. A memória compartilhada pode ser criada de duas maneiras:

* Eletronicamente (*Hardware*): o compartilhamento é feito em uma grande área de memória RAM que pode ser acessada por vários processadores, geralmente utilizada em supercomputadores;
* Logicamente (Software): aqui o compartilhamento é feito utilizando-se estruturas de dados especiais nos programas, para que possam realizar a troca de informação durante a sua execução.

### 4. Organização da Memória

A memória geralmente é dividida em três partes:

* Memória Cache: esta memória fica dentro do processador, é muito menor que a Memória Principal, mas possui uma importância muito grande, pois é utilizada para acelerar a obtenção de informações para o processador;
* Memória Principal: também chamada de Memória RAM, é ela que armazena os programas e informações que estão em uso no computador;
* Memória Secundária: Nesta memória ficam os dados temporariamente não utilizados, para liberar espaço na Memória Principal, e aqui também ficam armazenados os arquivos do usuário.

### 5. Coletor de Lixo

Uma última tarefa sob responsabilidade do Gerenciador de Memória é fazer a “coleta de lixo”. Esta tarefa consiste em coletar os dados não mais utilizados na memória, e liberar espaço, para poder ser utilizar-se de outros programas. Os sistemas operacionais modernos, como o MacOS X e MS-Windows XP, possuem esse mecanismo. Versões anteriores desses sistemas que não possuíam o Coletor de Lixo não realizavam a limpeza automática da memória. Se um programa não limpasse a área da memória que ele mesmo utilizou o sistema operacional não iria detectar que a memória poderia ser utilizada, o que acarretava na perda de recurso durante a utilização do computador (dependendo do tempo de uso contínuo do computador, nitidamente percebia-se uma queda no seu desempenho).

## Memória Virtual

### 1. Introdução

A memória virtual é um recurso utilizado na computação que cria para os programas um ambiente de memória contínua e na prática é construído em um ambiente de memória fragmentada, podendo até ser utilizados dispositivos de armazenamento para o salvamento de informações de maneira temporária. A utilização de memória virtual torna o uso da memória principal mais eficiente.

O entendimento de Memória Virtual deve ir além do conceito de “memória temporária armazenada em disco”. O aumento da memória é o primeiro ganho que se tem ao utilizar-se memória virtual. O conceito real de memória virtual é: “gerenciar a memória de forma que os programas em execução no computador tenham um ambiente de memória unificada e contínua”. Os principais sistemas operacionais em operação atualmente fazem uso da técnica de Memória Virtual para obter com isso um aumento no desempenho do computador.

A técnica de memória virtual foi inicialmente desenvolvida nos anos de 1959 a 1962, na Universidade Manchester, Inglaterra, para o computador Atlas. Antes da criação da memória virtual gerenciada pelo sistema operacional, os próprios programas eram responsáveis por realizarem o gerenciamento da informação, utilizando a memória principal e os dispositivos de armazenamento de acordo com a necessidade inerente ao próprio programa.

A principal motivação para a criação da memória virtual foi o pensamento de se proporcionar aos desenvolvedores de programas um recurso mais facilitado de utilização da memória do computador. Na década de 1960 começaram a surgir os primeiros modelos de computadores comerciais que possuíam o recurso de memória virtual disponível, mas como toda tecnologia recém lançada, passou por um processo de popularização difícil.

Os primeiros sistemas operacionais implementavam técnicas rudimentares de memória virtual, que em comparação com os sistemas proprietários, em que os programas gerenciavam sua própria memória, tinham um desempenho abaixo do desejado para justificar uma substituição. Foi a IBM no final da década de 1960 que oficializou a adoção da memória virtual nos computadores ao lançar alguns modelos que utilizavam esse recurso. A partir de então, todas as empresas passaram a implantar nos sistemas operacionais a capacidade de gerenciamento da memória através da Memória Virtual.

A utilização de memória virtual em um sistema computadorizado depende principalmente da arquitetura do computador, mais especificamente, do modo de endereçamento de memória utilizado pelo processador. A Intel, desenvolvedora dos processadores Pentium, adotou o recurso de memória virtual em seus processadores em 1982 com o lançamento do modelo 80286. A Microsoft passou a utilizar o recurso de memória virtual apenas em 1990, no lançamento do MS-Windows 3.

Os programas utilizam a memória reservando espaços de dados para serem armazenados o código do programa em si e os dados que serão processados pelo programa. Este espaço de dados é chamado de Espaço de Endereçamento. Quando um programa é iniciado, o próprio sistema operacional reserva um Espaço de Endereçamento para armazenar o código executável do programa, e à medida que o programa é utilizado pelo usuário, mais espaços de endereçamento podem ser solicitados para armazenar os dados que estiverem sendo informados.

Para o sistema operacional gerenciar de forma mais eficiente a memória do computador, os espaços de endereçamento que um programa está utilizando são agrupados em um bloco único (a Área de Armazenamento formada pelos Espaços de Endereçamento) que é manipulado e administrado como uma única seqüência de dados na memória.

### 2. Implementação da Memória Virtual

A maioria das implementações de gerenciadores de memória virtual divide a memória em Espaços de Endereçamento de pelo menos 2 KB de tamanho, este espaço de dados é chamado de Página de Memória. Cada página é uma área contínua de dados, não fragmentada. Quanto maior a quantidade de memória disponível no computador, maior será o tamanho de uma Página de Memória.

Para realizar o mapeamento entre o Endereço Físico de Memória e o Endereço de Memória Virtual é utilizada uma Tabela de Paginação. Cada registro na Tabela de Paginação contém: o endereço da página inicial na Memória Virtual e o endereço real na memória física (que pode ser um valor na memória principal ou na memória secundária). Pode existir uma única Tabela de Paginação para todo o sistema ou podem existir tabelas específicas para cada aplicação em uso no computador.

A Paginação é a técnica empregada para se realizar a transferência de dados inativos, que podem ser programas ou informações não utilizadas, retirando-os da memória principal e colocando-os na memória secundária. Para executar a paginação, a maioria dos sistemas operacionais atuais utiliza o disco-rígido como memória secundária. Para os programas, a soma da memória principal e memória secundária será a memória disponível para armazenar informações.

Quando um programa não está em uso, ou quando uma informação não está sendo acessada, a área de memória que está reservada para guardar o código do programa ou os dados da informação, será retirada da memória física principal e colocada na memória física secundária. O endereçamento virtual não será afetado.

Quando o programa voltar a ser utilizado ou quando a informação for acessada, a paginação será feita no sentido contrário, retirando o que está na memória secundária e colocando na memória principal. O Espaço Físico utilizado na memória principal não é necessariamente o mesmo utilizado anteriormente, o que precisa ser mantido é o Espaço Virtual de Endereçamento.

### 3. Páginas Permanentes

Algumas páginas armazenadas na memória virtual não podem, em qualquer hipótese, sofrer paginação. Os programas ou informações contidas nessas páginas devem obrigatoriamente estar disponíveis durante todo o tempo de utilização do computador, alguns exemplos são:

* A Tabela de Paginação, propriamente dita, é uma informação que deve estar sempre disponível na memória principal;
* Rotinas ativadas por Interrupção também devem estar sempre armazenadas na memória principal;
* Dispositivos que utilizam o recurso de Acesso Direto à Memória (*Direct Memory Access*, DMA) devem ter sempre disponível no mesmo espaço físico de memória as informações que estão sendo manipuladas. Como a paginação não garante que os dados retornarão ao mesmo endereço físico, as informações desses dispositivos não poderão ser paginadas.

## Memória Paginada

### 1. Introdução

Memória Paginada é a estratégia utilizada pelo sistema operacional para melhor gerenciar a memória do computador. Tanto a memória principal quanto a memória secundária será dividida em Páginas de Memória, formando a Memória Virtual. O conteúdo armazenado nas páginas de memória estará armazenado na memória principal ou na secundária e a alteração de uma página entre a memória principal e secundária é chamada de paginação.

A principal vantagem alcançada pela utilização da Memória Paginada é a facilidade na recuperação de informações que estão em páginas na memória secundária. A recuperação acontecerá por dois motivos: o processador estará carregando o código executável de um programa (o processo passou do estado de Bloqueio para o estado de Espera), ou um programa estará acessando alguma informação na memória virtual.

O sistema operacional, por meio do gerenciador de memória, detecta que essa informação não está na memória principal (essa situação é chamada de “Falha de Página”, ou “Page Fault” em inglês). Sempre que acontecer uma falha de página, o sistema operacional deverá realizar o seguinte procedimento:

* Encontrar a página de memória na Memória Secundária;
* Determinar qual área da Memória Principal será utilizada para receber os dados da

Memória Secundária, e armazenar esses dados na Memória principal;

* Passar para o programa a nova página de memória correspondente à informação solicitada.

O item dois pode ser expandido, caso não haja uma página de memória livre na memória principal (toda a memória principal está em uso) para armazenar os dados da memória secundária. Nesse caso, haverá duas paginações: uma retirando uma página da memória principal para a memória secundária, e outra no caminho inverso. Diz-se então que houve uma Troca de Páginas (*Page Swap*, em inglês).

### 2. Troca de Página

Muitos sistemas operacionais realizam a Troca de Página utilizando um algoritmo que verificará qual a página de memória menos utilizada. As páginas de memória menos utilizadas são consideradas como candidatas a uma eventual troca de página, pois a probabilidade de serem utilizadas logo após a troca é considerada pequena ou menor que a probabilidade de uso de uma página de memória que está sendo frequentemente acessada. Por exemplo, quando se está escrevendo um documento em um programa como o MS-Word ou OpenOffice Writer, as páginas de memória mais acessadas são as que possuem as folhas de texto mais próximas do ponto em que se está escrevendo o documento. Quando o documento atinge uma determinada quantidade de folhas, as primeiras folhas passam a ser menos acessadas, e consequentemente, as respectivas páginas de memória tornam-se então páginas menos utilizadas.

A maioria dos programas, quando estão em execução, assume um comportamento padrão no que diz respeito à utilização de memória, pois a frequência com que a memória é solicitada é constante. Este comportamento define uma característica importante no gerenciamento da memória virtual: o conjunto de páginas de memória mais utilizadas pelo programa. O programa irá ser executado de maneira mais eficiente se este conjunto estiver armazenado na memória principal do computador. Um programa que, em relação à quantidade total de memória principal, trabalhe com uma quantidade reduzida de informação e instruções e que esteja constantemente em uso, ficará sempre na memória principal.

Se houver um programa que tenha uma quantidade de instruções muito grande, e trabalhe com uma quantidade de informação maior que o espaço disponível na memória principal (por exemplo, um programa de edição de vídeo está sendo utilizado para editar um filme, trabalhará com uma quantidade de informação superior a 4 GB), a realização de Trocas de Páginas será tão grande que prejudicará o desempenho do sistema, pois os principais recursos do computador: processador, memória principal e memória secundária, estarão em uso realizando as trocas de páginas de memória.

Para minimizar o problema de excesso de Troca de Página, duas medidas podem ser tomadas:

* Aumentar a quantidade de memória principal: quanto maior for a memória principal do computador, maior será a quantidade de páginas de memória disponíveis dentro da memória principal;
* Diminuir a quantidade de programas em execução no computador: quanto menos programas estiverem em execução, menor será a necessidade por páginas de memória, aumentando a possibilidade de existirem páginas disponíveis.

A medida escolhida dependerá do tipo de utilização do computador e dos recursos financeiros disponíveis.

### 3. Diferença entre Paginação e Troca de Página

A Paginação, vista na Unidade 12, é a transferência de dados inativos da memória principal para a memória secundária, enquanto que a Troca de Página acontece quando uma página da memória secundária irá ocupar o espaço de uma página em uso na memória principal.

A maioria dos sistemas operacionais modernos implementam essas duas técnicas, mas fazem um tratamento diferenciado quanto à nomenclatura utilizada:

No Linux, o mecanismo de Troca de Página é unificado com a Paginação, não havendo separação conceitual. A diferenciação é feita apenas na direção da troca:

* Entrada de Página: termo utilizado para designar a transferência de uma página da memória secundária para a memória principal, *Page-In* em inglês.
* Saída de Página: termo utilizado para designar a transferência de uma página da memória principal para a memória secundária, *Page-Out* em inglês.
* No MS-Windows é criado um arquivo físico no disco para armazenar as páginas menos utilizadas, este arquivo é chamado de Arquivo de Paginação, pagefile.sys, e tanto a paginação quanto a troca de página são tratadas da mesma maneira.

### 4. Diferenças na Troca de Página entre o MS-Windows e o Linux

Nos sistemas operacionais MS-Windows, a Troca de Página é feita utilizando-se o arquivo pagefile.sys. A localização desse arquivo pode ser definida pelo usuário, mas geralmente fica no diretório principal do disco onde está instalado o sistema operacional, geralmente “C:\pagefile.sys”. O tamanho deste arquivo também pode ser definido pelo usuário, podendo ser de dois tipos:

* Tamanho Variável: à medida que mais páginas são necessárias, o arquivo de paginação irá ser incrementado.
* Tamanho Fixo: um tamanho fixo é definido e não será alterado.

A vantagem de ter o arquivo de paginação com tamanho variável é a garantia de que sempre espaço disponível para armazenar os programas e seus dados (limitado pelo tamanho total da unidade de armazenamento). Mas isso gera o inconveniente de acarretar em uma possível fragmentação do arquivo pagefile.sys, que implicará em perda do desempenho do computador. A utilização de um arquivo com tamanho fixo é uma garantia de que o arquivo jamais sofrerá fragmentação (contanto que o arquivo seja criado logo após a instalação do sistema operacional), mas limita a quantidade de memória disponível no computador. Para contornar esse problema, pode-se definir um arquivo com tamanho grande, mas pode-se comprometer o espaço disponível na unidade de armazenamento.

No Linux uma estratégia diferente é adotada: o arquivo de paginação fica completamente separado em uma Partição. Uma partição é uma divisão funcional da unidade de armazenamento. Se uma unidade de armazenamento for particionada em duas unidades, cada unidade será gerenciada pelo sistema operacional de maneira independente da outra unidade. O que acontece na prática é que se tem um arquivo de paginação fixo no Linux.

## Memória Segmentada

### 1. Introdução

A Segmentação da Memória consiste na divisão da memória em segmentos de dados. Cada segmento possuirá as seguintes propriedades: a sua localização física na memória, permissão de acesso, tamanho e um identificador. Se um processo possuir permissão para utilizar um segmento, e se o processo solicitar acesso a uma informação dentro do segmento - poderá fazer isso, caso contrário um aviso de erro será passado ao processo (e consequentemente ao usuário).

Além das propriedades citadas, dependendo do sistema operacional, o segmento pode possuir um indicador informando se sua localização está na memória principal ou na memória secundária. Caso esteja na memória secundária, e seja acessado por algum processo, o sistema operacional irá realizar o carregamento do segmento na memória principal para poder acessar as informações contidas nele.

### 2. Unidade de Gerenciamento da Memória

A segmentação da memória só será possível se houver no computador, dentro do próprio processador ou como um circuito eletrônico independente, uma Unidade de Gerenciamento da Memória (*Memory Management Unit*, MMU). Esta unidade será responsável por fazer o mapeamento entre os endereços físicos e os endereços dos segmentos de memória. Os processos não acessam diretamente as áreas de memória do computador, em vez disso, cada processo ao ser iniciado será associado a um segmento de memória específico. Quando um processo realiza o acesso a uma determinada área de memória, dentro do seu segmento, a Unidade de Gerenciamento da Memória irá acessar a área física correspondente ao segmento designado ao processo. Para fazer isso, a Unidade de Gerenciamento irá executar as seguintes verificações:

* O identificador do segmento do processo;
* A localização na memória física do segmento, isto é, qual o endereço inicial do segmento na memória;
* A localização área de memória acessada pelo processo;  A quantidade de informação solicitada pelo processo.

Se o processo em questão fizer a solicitação corretamente, isto é, solicitar acesso a uma área de memória dentro do seu segmento, a Unidade de Gerenciamento da Memória irá completar essa operação e passar ao processo os dados contidos na área de memória que foi solicitada. Se por acaso o processo fizer uma solicitação que não seja permitida, por exemplo, o processo solicita acesso a uma área de memória dentro do segmento a que pertence, mas que ultrapasse o tamanho do segmento, essa solicitação não será atendida, e um aviso de erro será passado ao processo.

Se na realização do acesso ao segmento, este não se encontrar na memória principal, o sistema operacional irá verificar se o segmento encontra-se armazenado na memória secundária. Encontrando o segmento, o mesmo será carregado na memória principal para que as informações contidas nele possam ser acessadas.

Se um processo tentar acessar uma área de um segmento de memória que oferece permissão para o processo em questão ou se o processo tentar acessar uma área de memória fora do seu segmento, o sistema operacional facilmente conseguirá identificar esse fato, pois cada segmento possui um identificador e a definição da permissão de acesso. Se o segmento permitir o acesso de vários processos diferentes, as informações contidas nele poderão ser acessadas por qualquer processo. Se o segmento permitir o acesso apenas a um processo (quem define o tipo de permissão é o primeiro processo criado associado ao segmento) a informação só poderá ser lida pelo processo que foi inicialmente associado ao segmento. Quando o processo finaliza a sua execução, o segmento de memória é liberado (ou pelo próprio programa ou pelo coletor de lixo do sistema operacional) e assim o segmento fica disponível para ser utilizado por outro processo.

### 3. Memória Virtual Segmentada

Em alguns sistemas, a memória virtual não utiliza páginas de memória, e sim segmentos de memória. Nesse caso, serão criados com a possibilidade de terem o seu tamanho alterado durante a utilização do computador.

A Unidade de Gerenciamento da Memória deve estar preparada para lidar com essa situação, pois em um determinado momento um processo pode solicitar acesso a uma área inicialmente indevida, mas esta mesma área pode ser adicionada ao segmento posteriormente, permitindo sua utilização.

### 4. Combinação da Segmentação e Paginação

Nos sistemas IBM System/38 (Mainframes) o gerenciamento da memória é feito utilizando-se a paginação e a segmentação. Cada segmento de memória é dividido em páginas de memória. Nos processadores Intel x86, os segmentos de memória também são divididos em páginas de memória e cada segmento é tratado como sendo uma memória virtual independente. Mas esse recurso só poderá ser utilizado se o sistema operacional em execução no computador permitir que se realize essa distinção, o que não acontece nos dias de hoje