

[Aula 16] Hierarquia de memória 3: *Memória virtual*

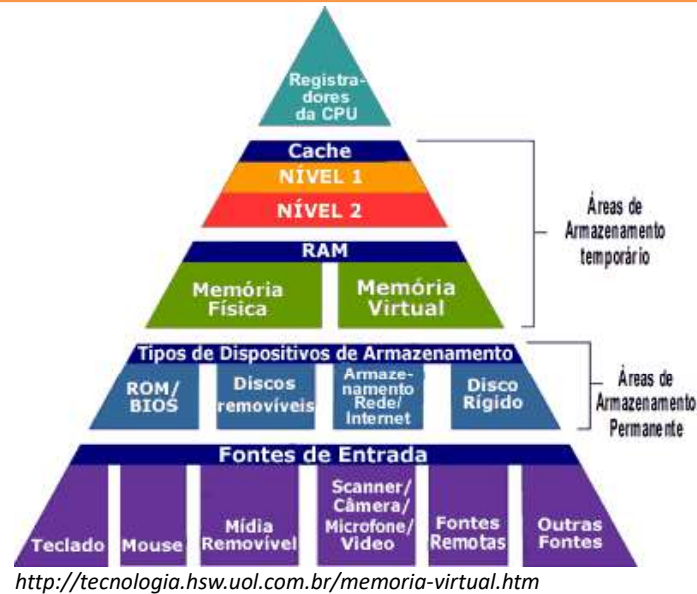
Prof. João F. Mari
joaof.mari@ufv.br

Aula 16 - Hierarquia de memória 3 - Memória virtual

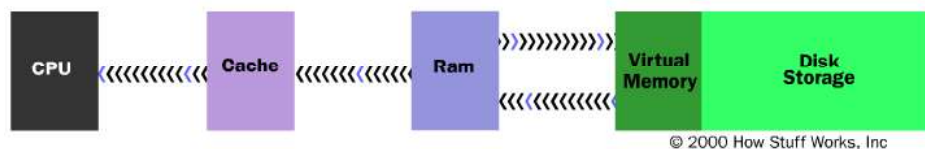
Roteiro

- **[REVISÃO]** Hierarquia e gerenciamento de memória
- Memória virtual
- **[EX]** Dois programas compartilhando a memória
- Memória virtual: escolhendo parâmetros
- Tradução de endereços
- Falhas de páginas (falhas)
- Políticas de substituição de páginas
- TLB – tornando a tradução de endereços mais rápida.
- **[EXERCÍCIOS]**

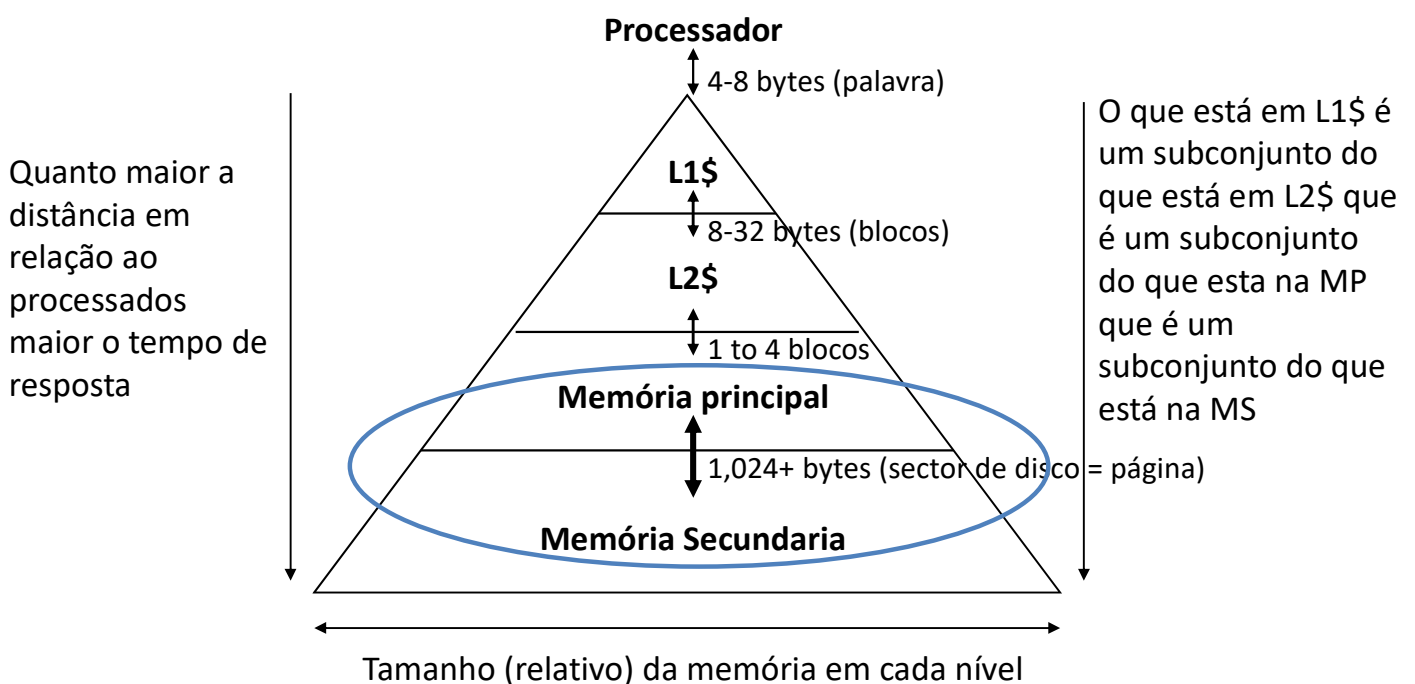
[Revisão] Hierarquia e gerenciamento de memória



Memory Management



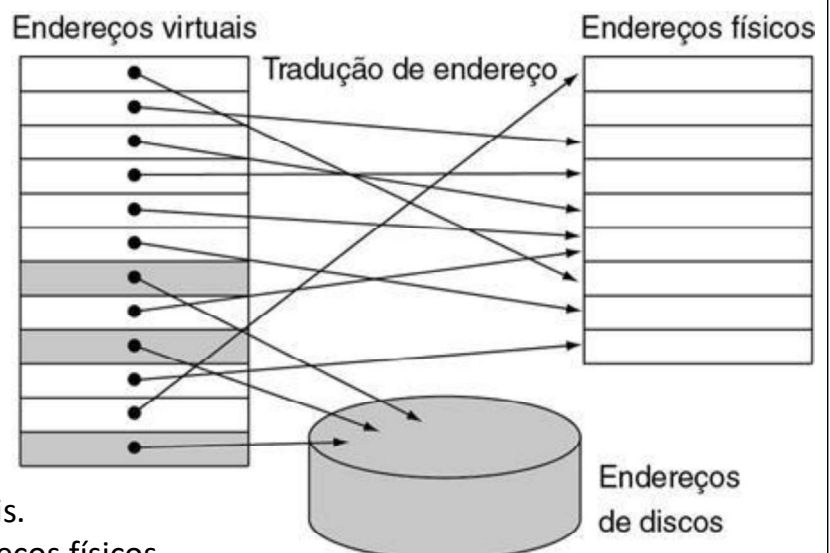
[Revisão] Hierarquia e gerenciamento de memória



Memória virtual

- Utilização da memória principal como cache para a memória secundária
 - **Permite o compartilhamento seguro e eficiente da memória entre vários programas.**
 - **Simplifica o carregamento de um programa em memória:**
 - Relocação de código: programas são compilados com endereços virtuais que podem ser mapeados em qualquer posição de memória.
 - **Permite que programas que sejam maiores que a memória física instalada sejam executados.**
- Princípio de localidade:
 - Um programa tende a utilizar uma porção pequena de seu espaço de endereçamento durante um período de tempo.
- Cada programa é compilado com o seu próprio espaço de endereçamento (espaço de endereçamento virtual);
 - Durante a execução, cada endereço virtual deve ser traduzido em um endereço físico (um endereço existente na memória principal).

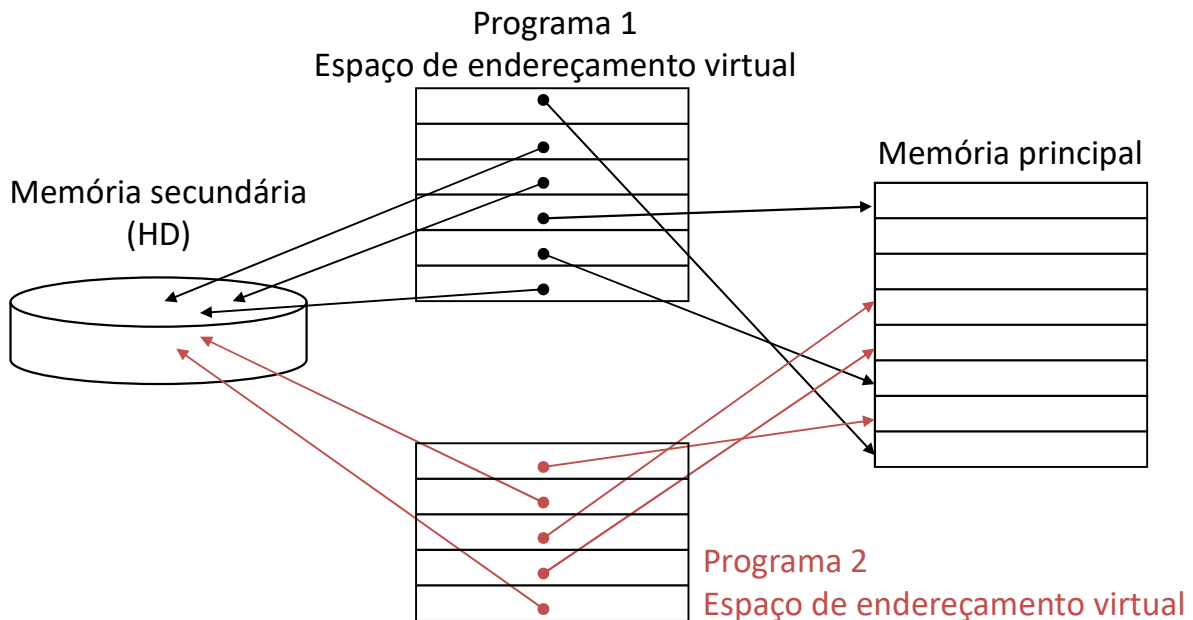
Memória virtual



- O processador gera endereços virtuais.
- A memória é acessada usando endereços físicos.
- Tanto memória virtual quanto memória física são organizadas em páginas (blocos de memória).
- Uma página virtual é realmente mapeada para um endereço físico.
- Quando uma página virtual está ausente da memória principal ela reside no disco.
- Mais de um endereço virtual pode apontar para um mesmo endereço físico. Permite que programas diferentes compartilhem dados ou códigos.

[EX] Dois programas compartilhando a memória

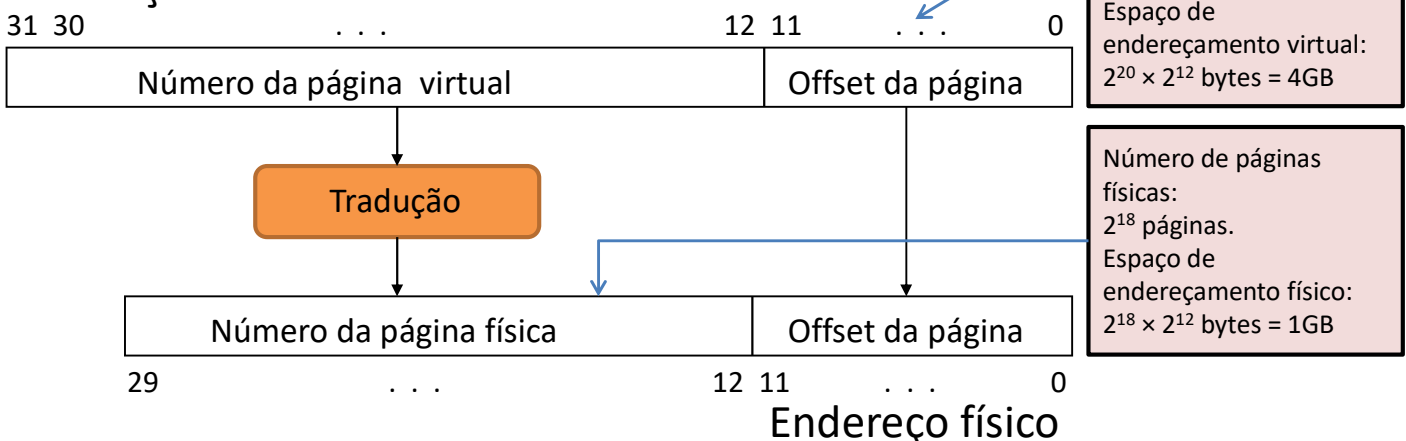
- O espaço de endereçamento do programa é dividido em páginas (todas de tamanho fixo) ou segmentos (tamanhos variáveis)
 - O endereço inicial de cada página (localizada na memória principal ou na memória secundária) está registrado na **tabela de páginas** do programa.



Memória virtual

- Um bloco de memória virtual é chamado de página
 - Uma falha de memória virtual é chamada de falta de página (*page fault*)
 - o processador produz um endereço virtual traduzido por uma combinação de hardware e software para um endereço físico: processo de mapeamento de endereço ou tradução de endereço
- $\text{Endereço virtual} = \text{número de página virtual} + \text{offset da página}$

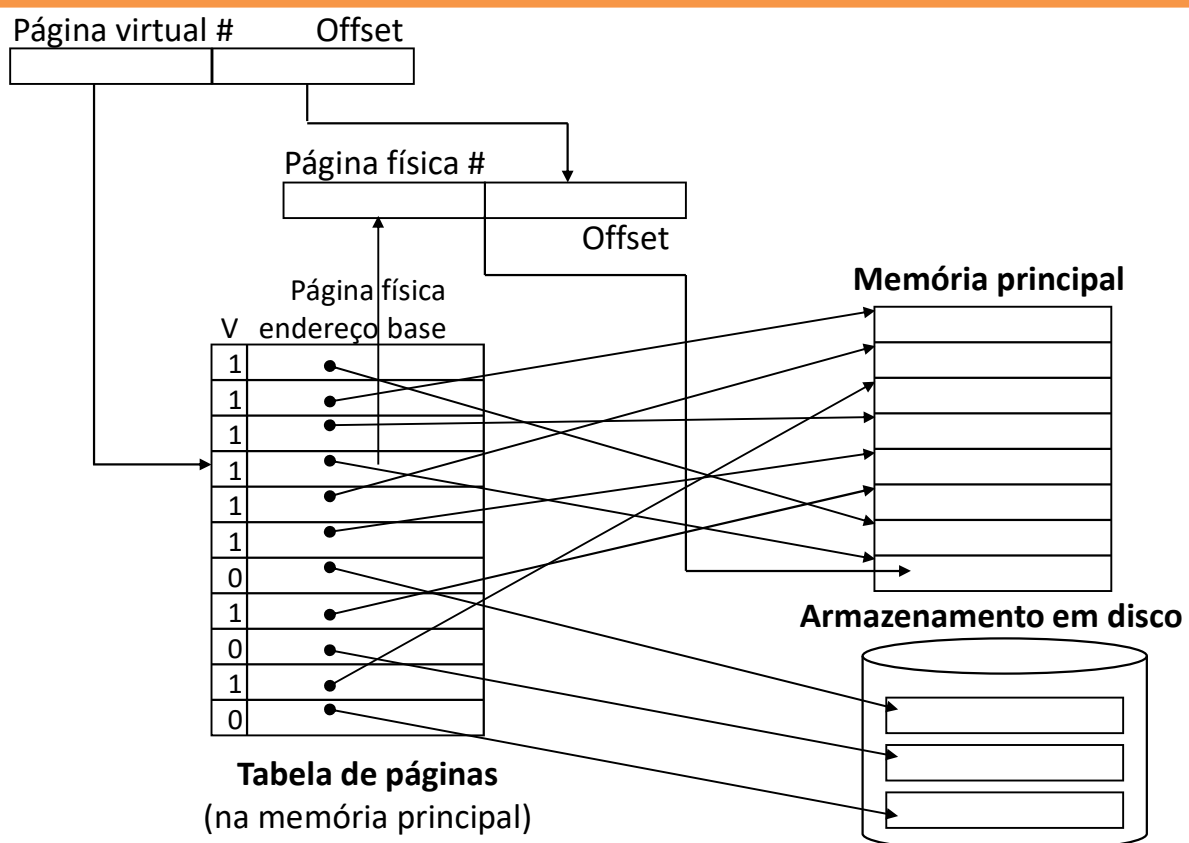
Endereço virtual



Memória virtual: escolher os parâmetros.

- Uma falha (falta de página) em sistemas de memória virtual possuem custo alto:
 - A penalidade de falha elevada;
 - Milhões de ciclos de *clock* para ser processada;
 - Decisões importantes nos projetos de sistemas de memória virtual.
- As páginas devem ser grandes o suficiente para tentar amortizar o tempo de acesso
 - Estações de trabalho (4KB, 16KB) e servidores (32KB, 64KB)
 - Sistemas embarcados (1KB)
 - Organizações que reduzem a taxa de faltas de página são atraentes:
 - **Principal técnica:** posicionamento totalmente associativo das páginas na memória.
 - As faltas de página podem ser tratadas em nível de software:
 - O *overhead* será pequeno se comparado ao alto tempo de acesso ao disco;
 - Reduções nas taxas de falhas compensam o custo dos algoritmos de gerenciamento.
 - O **write-through** não funcionará para a memória virtual
 - Escritas levam muito tempo;
 - Sistemas de memória virtual utilizam **write-back**.

Traduções de endereços



Faltas de página

- Falta de página:
 - Bit de validade (V) para uma página virtual desligado:
 - O sistema operacional recebe o controle por meio de um mecanismo de exceção.
 - Ele encontra a página no próximo nível da hierarquia, geralmente o disco magnético; e
 - Decidir onde colocar a página requisitada na memória principal (paginação).
- O endereço virtual por si só não diz imediatamente onde está a página no disco:
 - O sistema operacional monitora o local no disco de cada página em um espaço de endereçamento virtual;
 - O sistema operacional cria um espaço no disco para todas as páginas de um processo no momento que o processo é criado, chamado de área de swap.
- **Tabela de páginas:**
 - Uma estrutura auxiliar para registrar onde cada página está armazenada no disco;
 - Pode ser parte da tabela de páginas ou uma estrutura auxiliar indexada da mesma maneira.

Faltas de página

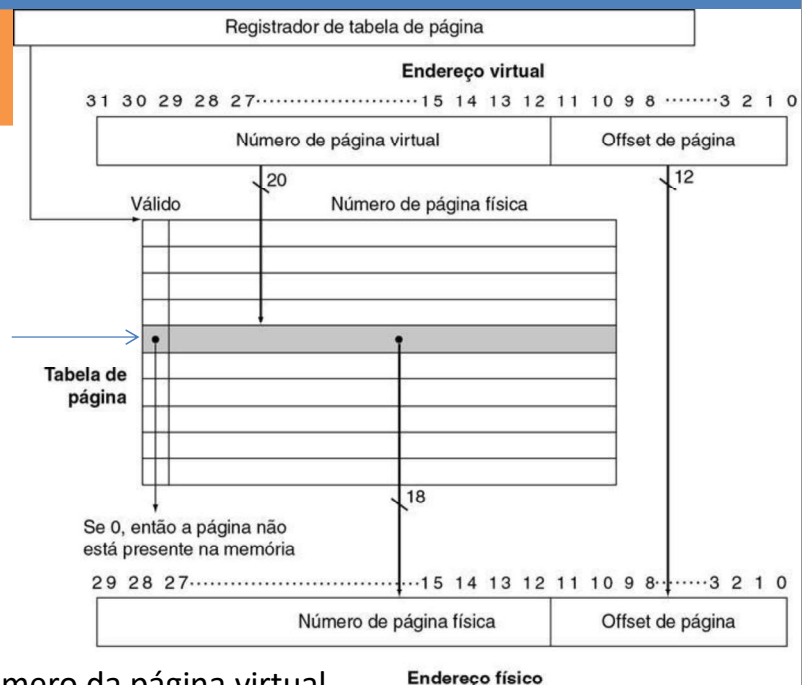
- Se (falta de página):
 - Consulta a tabela de páginas;
 - Transfere a página que faltou do disco para a memória;
 - Memória cheia?
 - Sim → Substitui uma página da memória
 - Usa uma política de substituição de páginas.
 - Não → Atualiza tabela de páginas.

Política de substituição de páginas

- O sistema operacional cria uma estrutura de dados que controla quais processos e quais endereços virtuais usam cada página física
- Quando ocorre uma falha:
 - Se todas as páginas na memória principal estiverem em uso, o sistema operacional precisa escolher uma página para substituir
 - Algoritmo LRU (*last recently used*)
 - Os sistemas operacionais monitoram as páginas e atualizam o bit de referência (ou bit de uso) que é ligado sempre que uma página é acessada

Faltas de página

Nesse exemplo, cada entrada da Tabela de páginas precisa de apenas 19 bits (18 + 1), porém, para facilitar a indexação ela é arredondada para 32 bits (4 bytes).

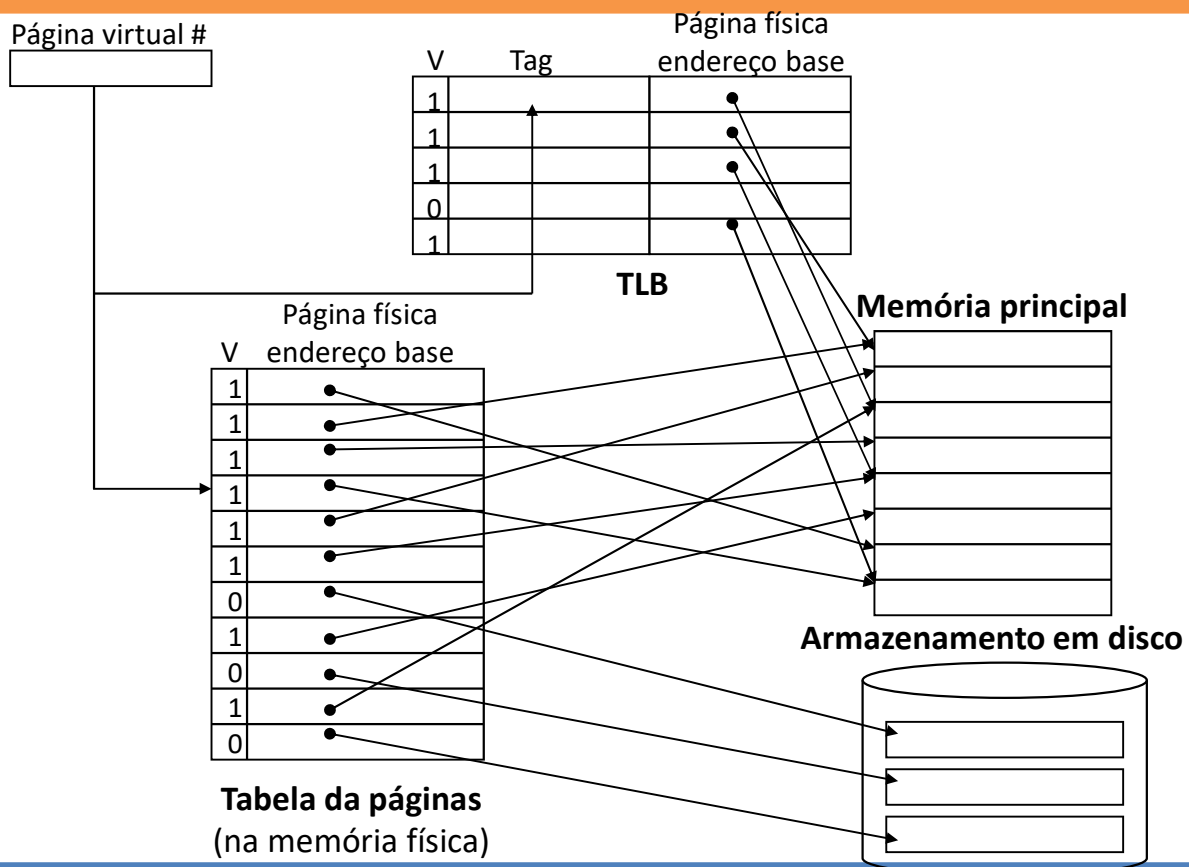


- A tabela de páginas é indexada pelo número da página virtual.
- O ponteiro da tabela de páginas indica o início da tabela na memória principal.
- O espaço de endereçamento virtual é 2^{32} bytes, ou 4GB.
- O espaço de endereçamento físico é 2^{30} bytes. Permite uma memória principal de até 1GB.
- O tamanho da página é 2^{12} bytes, ou 4096 bytes, ou ainda 4KB.
- O número de entradas na tabela de páginas é 2^{20} .
- O tamanho (em bytes) dessa tabela de páginas é $2^{20} * 4\text{bytes} = 4\text{MB}$ para cada processo.

Tomando a tradução de endereços mais rápida: a TLB

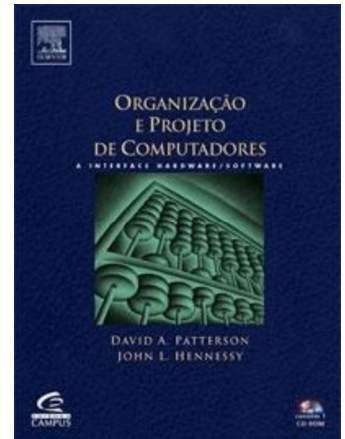
- As tabelas de páginas são armazenadas na memória principal
 - Cada acesso à memória por um programa pode levar, no mínimo, o dobro do tempo:
 - Um acesso para obter o endereço físico e um acesso para obter os dados
 - Quando uma tradução para uma página é realizada, possivelmente essa tradução será repetida em um futuro próximo (princípio da localidade temporal)
 - Processadores modernos incluem uma cache especial que controla as traduções usadas recentemente:
- TLB (*translation-lookaside buffer*):**
 - Cache que monitora os mapeamentos de endereços recentemente usados para evitar acessos à tabela de páginas.

Faltas de página



Bibliografia

1. PATTERSON, D.A; HENNESSY, J.L. **Organização e Projeto de Computadores: A Interface Hardware/Software**. 3a. Ed. Elsevier, 2005.
 - Capítulo 7.
2. Notas de aula do prof. Luciano J. Senger:
 - <http://www.ljsenger.net/classroom.html>
3. Notas de aula da profa. Mary Jane Irwin
 - CSE 331 Computer Organization and Design
 - <http://www.cse.psu.edu/research/mdl/mji/mjicourses>



FIM

- FIM:
 - **Aula 16** – Hierarquia de memória 3 – Memória virtual.
- Próxima aula:
 - **Aula 17** – Entrada e Saída