

Glossário Técnico

Módulo II: A Memória Cache (L1, L2 e L3)

Explore os conceitos fundamentais da memória cache e sua hierarquia de níveis, essenciais para compreender o desempenho dos processadores modernos.

Cache e Hierarquia

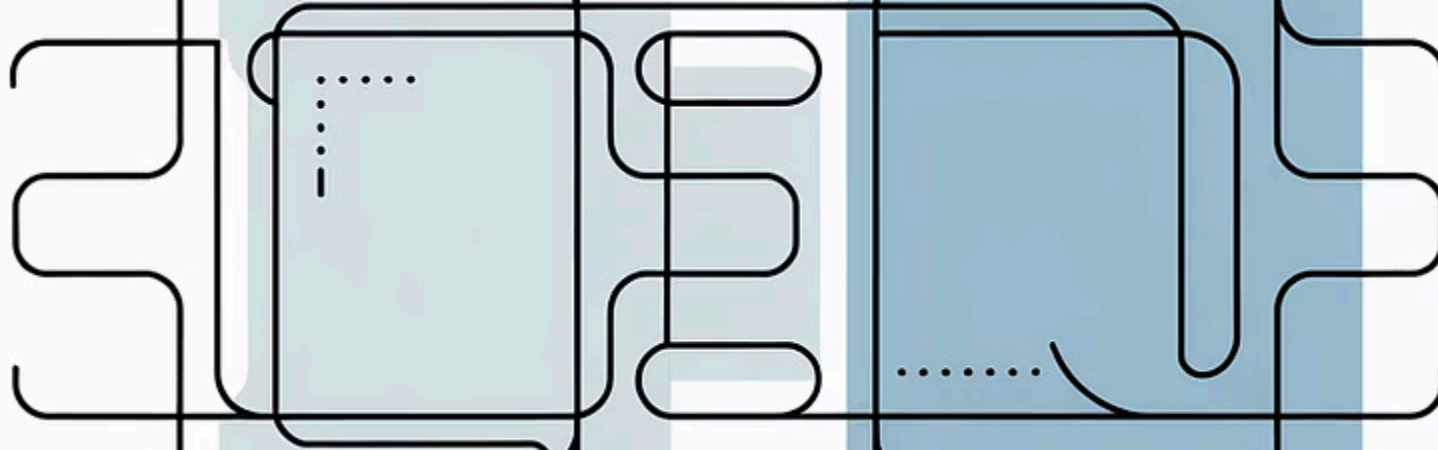
Cache

Memória ultrarrápida dentro do processador que armazena temporariamente dados e instruções frequentes, reduzindo o tempo de acesso à RAM.

Hierarquia de Cache

Estrutura em níveis (L1, L2 e L3) organizada por velocidade e capacidade. L1 é a mais rápida, L3 a mais ampla, L2 atua como intermediária.





Cache Hit vs Cache Miss

Cache Hit ✓

Situação em que o dado procurado já está armazenado na cache. O acesso é imediato e ocorre em poucos ciclos de clock.

Cache Miss ☒

Quando o dado não está na cache, obrigando o processador a buscá-lo na RAM, com aumento significativo da latência.

📄 **Latência:** Tempo necessário para concluir uma operação de leitura ou escrita, medido em ciclos de clock.

Os Três Níveis de Cache



L1 Cache (Nível 1)

Camada mais próxima do núcleo. Extremamente rápida e pequena (16 KB a 128 KB). Divide-se em L1I (instruções) e L1D (dados).



L2 Cache (Nível 2)

Camada intermediária com tamanho maior (512 KB a 2 MB por núcleo). Mantém dados recentes e atua entre L1 e L3.



L3 Cache (Nível 3)

Camada compartilhada entre todos os núcleos. Grande (4 MB a 64 MB), otimiza comunicação entre núcleos e reduz tráfego com a RAM.

Técnicas de Otimização



Prefetching

A CPU antecipa a necessidade de dados e os carrega na cache antes de serem requisitados, reduzindo cache misses.



MESI Protocol

Protocolo de coerência (Modified, Exclusive, Shared, Invalid) que garante dados atualizados e consistentes em sistemas multinúcleo.



Cache Coherence

Mecanismo que assegura que diferentes núcleos tenham a mesma visão dos dados, evitando inconsistências.

Conceitos Avançados



Associatividade de Cache

Estratégia que define como blocos de dados são armazenados. Uma cache "2-way associative" permite dois locais possíveis por bloco, aumentando chances de cache hit.



Pipeline

Conjunto de estágios de execução que processam várias instruções simultaneamente. Seu desempenho depende da rapidez da cache.



Smart Cache

Caches dinâmicas e inteligentes que redistribuem espaço e prioridade adaptativamente entre núcleos conforme a carga de trabalho.

Métricas de Desempenho

Throughput

Taxa efetiva de processamento de dados. Uma cache eficiente aumenta o throughput, permitindo mais instruções em menos tempo.

Thermal Design Power (TDP)

Caches maiores consomem mais energia e geram mais calor, afetando o equilíbrio térmico e energético do processador.



Políticas de Escrita

Write-Through

Os dados são gravados simultaneamente na cache e na RAM, garantindo consistência total entre os níveis de memória.

Write-Back

Os dados são gravados primeiro na cache e enviados para a RAM apenas quando substituídos, otimizando o desempenho geral.

Ambas as políticas têm vantagens específicas: write-through prioriza integridade, enquanto write-back maximiza velocidade.