

SISTEMI OPERATIVI

Evoluzione architetture
Prof. Viglietti Francesco
Classe 4 B info
A.S. 2010-11

Architettura 1

L'architettura della CPU è il modo in cui le parti sono state progettate per lavorare insieme

Le architetture possono classificarsi in base al grado di parallelismo che permettono:

- SISD: 1 sola CPU che esegue un'istruzione per volta (macchina di Von Neumann)
- MISD: può eseguire più istruzioni sui propri dati, mentre si esegue un'istruzione si iniziano ad eseguire le altre (Pipeline)
- SIMD: può eseguire la stessa istruzione contemporaneamente su più dati (per multimedia)
- MIMD: relativa ai sistemi multiprocessore

Architettura 2

SISD → l'esecuzione del processo è sequenziale ed ha le seguenti fasi:

- › Fetch: caricamento dell'istruzione dalla memoria, viene individuata mediante il registro *Program Counter*.
- › Decode: decodifica dell'istruzione, viene associato il codice da eseguire e aggiornato il *PC*.
- › Execute: esecuzione dell'istruzione.

MISD → Il processore viene suddiviso in più unità (pipelining) ad esempio una per ogni fase. La velocizzazione sta nel fatto che tutte le unità possono lavorare contemporaneamente. Un ulteriore miglioramento consiste nell'aumentare ancora il numero delle fasi (aggiungendo lettura/esecuzione/scrittura in memoria) e quindi della pipeline.

Istruzioni e Gestione della memoria

In base alla lunghezza di istruzioni si hanno architetture diverse: CISC (complex instruction set computing) → hanno molti tipi di istruzioni con differente lunghezza.

RISC (reduced instruction set computing) → poche istruzioni di lunghezza fissa

EPIC (explicitly parallel instruction computing) → usa istruzioni di lunghezza fissa combinate insieme in un'istruzione più lunga

La memoria può essere gestita con un'organizzazione:

Lineare → sequenza di locazioni accessibili mediante indirizzo lineare

Segmentata → suddivisa in segmenti dal programmatore, accesso mediante indirizzo iniziale del segmento sommato all'offset

La cache

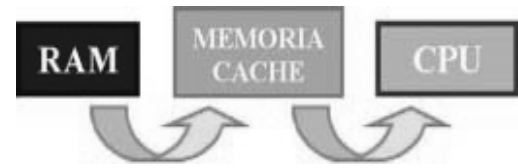
La cache è una memoria molto veloce posta tra CPU e memoria centrale. Serve per aumentare le prestazioni del sistema, ed è trasparente alla CPU.

Si hanno generalmente 3 livelli di cache:

L1 → integrata nel chip della CPU

L2 → esterna al chip, ma collegata direttamente alla CPU mediante il Bus Back Side che lavora alla stessa frequenza

L3 → esterna, sulla M/B



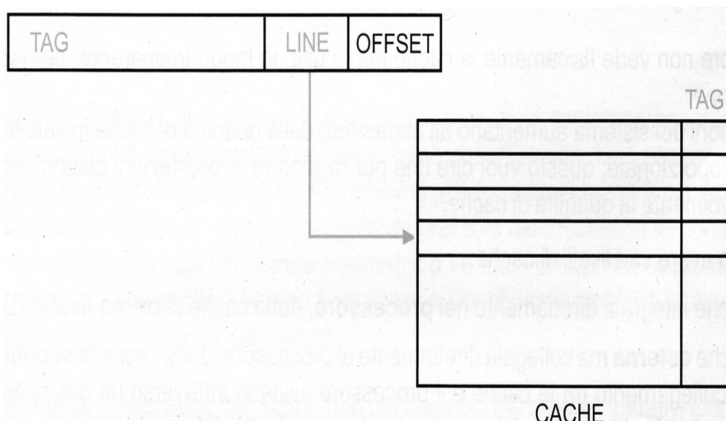
La cache è organizzata in blocchi (linee). Ad ogni operazione di lettura le informazioni vengono cercate prima in cache, e solo successivamente in RAM. Ad ogni operazione di scrittura la cache può usarsi come:

Write-through → scrittura sia in cache che in RAM

Write-back → scrittura solo in cache, solo su sostituzione viene scritta in RAM

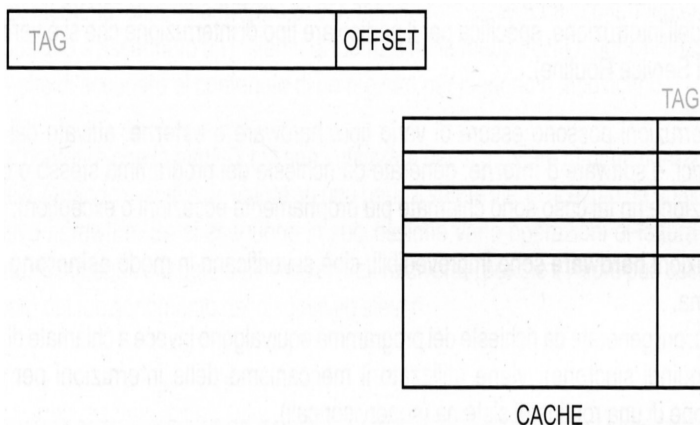
Metodi di gestione della cache 1

Metodo Diretto → ogni blocco di RAM viene allocato in un preciso blocco di Cache. L'indirizzo di memoria è diviso in 3 parti, **LINE** indica la linea di cache dove è allocato il blocco. Tutti i blocchi con **TAG** differenti vengono allocati nella stessa linea, ed identifica quale blocco sta in cache. **L'OFFSET** indica la posizione dell'informazione nella linea



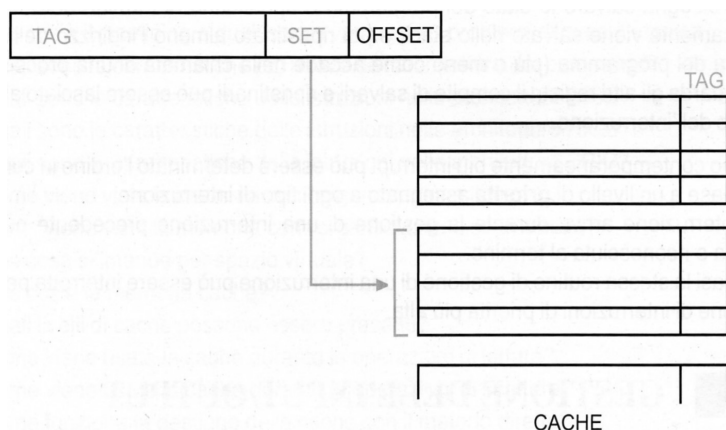
Metodi di gestione della cache 2

Metodo completamente Associativo → ogni blocco di RAM viene allocato in un qualsiasi linea della Cache. L'indirizzo di memoria è diviso in 2 parti, il **TAG** serve per trovare il blocco nella cache. Ricerca sequenziale, svantaggi e vantaggi.



Metodi di gestione della cache 3

Metodo Associativo a n vie → Cache divisa in gruppi di n linee. L'indirizzo di memoria è diviso in 3 parti, SET indica il gruppo in cache dove è allocato il blocco. Nel blocco vengono confrontati i TAG fino a trovarlo o ad esaurimento del gruppo. Se non c'è allora bisogna andare ad allocarlo, e se il gruppo non ha linee libere, bisogna rimpiazzare la linea di un determinato blocco secondo una determinata strategia.
















Interrupt

Un'interruzione è una particolare situazione, determinata dall'HW (esterne ed asincrone) o dal SW (interne chiamate eccezioni), che richiede un'intervento immediato. Si deve quindi attivare la routine di gestione delle interruzioni. Alcune interruzioni possono essere mascherate (alla CPU). Al momento dell'interrupt bisogna fare in modo che si salvi lo stato di esecuzione del programma ed il PC, in modo da non riprendere l'esecuzione del programma dopo l'interrupt.

Bisogna anche prevedere un meccanismo di priorità, anche nel caso di più interrupt contemporanei

Evoluzione processori intel

	N transistor	Frequenza	Processo produttivo [µm]	bit	N core
 4004	2.300	108 kHz	10,00	4	1
 8008	3.500	800 kHz	10,00	8	1
 8086	29.000	4.7 - 10 MHz	3,00	16	1
 386	134.000	16 - 40 MHz	1,50 - 1,00	32	1
 486	275.000	16 - 133 MHz	0,80	32	1
 Pentium	3.100.000	60 - 300 MHz	0,80 - 0,25	32	1
 Pentium II	7.500.000	233 - 450 MHz	0,35 - 0,25	32	1
 Pentium III	9.5 - 46 mil	450 - 1400 MHz	0,25 - 0,13	32	1
 Pentium IV	42 - 178 mil	1.3 - 3.8 GHz	0,18 - 0,065	32 - 64	1
 Pentium D	230 - 376 mil	2.66 - 3.6 GHz	0,09 - 0,065	64	2
 Core 2 Duo	167 - 410 mil	1.06 - 3.33 GHz	0,065 - 0,045	64	2
 Core 2 Quad	582 - 820 mil	2.4 - 3 GHz	0,065 - 0,045	64	4
 Core i7	731 mil	2.66 - 3.2 GHz	0,045	64	4