

Valutazione delle prestazioni

Un elaboratore può essere classificato secondo diversi parametri, in gran parte dipendenti dal dominio applicativo. Questi parametri possono essere la potenza di calcolo (IBM Roadrunner system - Potenza 1100 TFlops), potenza grafica, transazioni gestite in un'unità di tempo (ad esempio Database), rapporto costo/prestazioni.

Per i server, invece, un parametro che ne guida il progetto architettonico è il MTBF (Mean Time Between Failures), da cui si può determinare un Costo orario di Disservizio, cioè la perdita di denaro che si subisce fintanto che il server non è nuovamente operativo.

Le prestazioni di un elaboratore però non dipendono solo dalla CAPU, ma sono anche conseguenza dell'architettura di sistema, dalle prestazioni e dalla tipologia dei periferici e dal software di base che lo gestisce.

A verifica di quanto detto è utile sapere che le prestazioni del processore non sono sempre quelle che hanno l'impatto maggiore sulle prestazioni di sistema. Ad esempio, in un sistema orientato ad applicazioni web il 40% del tempo è svolto in esecuzioni di istruzione e il restante 60% in operazioni di I/O. Se per caso venisse sostituito il processore con uno 10 volte più veloce otterremmo un miglioramento globale di:

$$SPEEDUP \text{ (prestazioni senza miglioramento / prestazioni con miglioramento)} \\ = T_{vecchio} / T_{nuovo} = T(0.6+0.4) / T(0.6/0.4/10)$$

$$SPEEDUP = 1/0.64 = 1.56$$

Pur migliorando le prestazioni della cpu di 10 volte, a malapena abbiamo ottenuto un miglioramento del 50%.

Un altro errore molto comune è quello di pensare che il costo dominante di un sistema sia quello costituito dal costo del processore. Infatti, nella maggior parte dei sistemi con rapporto costo/prestazioni migliore, il costo maggiore appartiene ai dischi e non alla cpu.

I principali parametri per la valutazione dello sviluppo e il miglioramento delle prestazioni sono:

- **Throughput** (ampiezza di banda) = (Operazioni, Azioni, Compiti) / Tempo, corrisponde alla quantità di operazioni svolte in un'unità di tempo.
- **Latenza**, ovvero il tempo impiegato per la completa esecuzione di un'operazione.

Consumo energetico

Uno dei problemi più rilevanti è la potenza dissipata sia a livello dell'elaboratore sia a livello del microprocessore (filosofia GREE IT).

Se pensiamo a un Pc, esso assorbe una potenza elettrica pari a 200W, e se pensiamo che operi per 2000h/anno, abbiamo un consumo di 400 KW/h. Se poi teniamo conto che in tutto il mondo sono attivi 300 milioni di PC, il consumo totale raggiunge i 120 TWh. Questa potenza consumata globale continua a crescere ogni anno di un valore uguale a circa l'8% - 10%.

Se ora confrontiamo queste cifre con l'ammontare di energia elettrica prodotta in Italia, che è pari a 350 Twh, capiamo come un piccolo risparmio del 10% sul consumo, possa farci risparmiare l'uso di alcune centrali.

Se invece pensiamo ai 10 primi supercomputer, essi hanno un consumo di 24.000 kW, con una potenza totale di calcolo di 5000 Tflops. Per capire le grandezze in gioco, 24000KW sono anche la potenza che viene consumata da 12 TGV.

La potenza dissipata viene definita come:

$$Potenza_{dissipata} = \frac{1}{2} * capacita_{carico} * tensione^2 * frequenza$$

Per esempio, quindi, basterebbe diminuire la tensione di alimentazione per avere una notevole riduzione del consumo.

Inoltre, l'energia dissipata (potenza per tempo) è un parametro essenziale per tutti i sistemi portatili alimentati a batteria (ne determina la durata).

Esempio: Alcuni microprocessori vengono progettati per funzionare con una tensione di alimentazione regolabile, così una diminuzione del 15% della tensione riduce anche la frequenza di clock.

$$Riduzione_{potenza} = \frac{((tensione + 0.85)^2 * frequenza * 0.85)}{((tensione^2) * frequenza)} = 0.61$$

Si ottiene significativa riduzione consumo a fronte di una riduzione delle prestazioni del solo 15%.

Ogni core di un processore possiede un'unità di controllo della temperatura della zona di silicio. Se questa supera una data soglia l'unità di controllo, l'unità di controllo riduce la frequenza di clock.

Classificazione

I microprocessori vengono classificati in 3 grandi gruppi:

- General Purpose: sono processori ad alte prestazioni e versatili, non limitati ad un solo possibile utilizzo. Fanno parte di questo gruppo i processori dei personal computer.
- Special Purpose: sono processori realizzati per compiti specifici. Ne fanno parte i microprocessori delle schede grafiche
- Embedded: sono sistemi a microprocessore progettati per una determinata applicazione, spesso con una piattaforma hardware ad hoc. Sono quindi integrati completamente nel sistema, di cui ne gestiscono tutte o quasi le funzionalità.

Ad oggi il mercato è costituito per il 98% da microprocessori per applicazioni embedded come elettrodomestici, periferiche computer e altro. I microprocessori per personal computer, pur rappresentando solo lo 0.2% del numero totale di pezzi prodotti, assorbono però la metà del fatturato totale.

Architettura dei microprocessori

Un microprocessore è composto da più unità:

- Unità di Controllo + ALU (Algebrical Logic Unit)
- Unità di Decodifica Istruzioni (determina la tempistica dovuta ai tempi di Fetch, Decode ed Execute. Nelle architetture CISC questi tre tempi sono variabili per la diversa lunghezza delle istruzioni, dal numero di microistruzioni necessarie per la loro implementazione e dal tipo di istruzione (1 o più argomenti, ecc...))
- Unità di Gestione Indirizzi
- Registri
- Unità di Gestione BUS

Sotto il punto di vista della tempistica il microprocessore può essere suddiviso sotto due gruppi. Uno che contiene 'ALU, i registri, l'unità di controllo, e l'altra che contiene l' interfaccia verso il BUS (Front-side Bus).

Questi due insiemi possiedono un clock differente, il primo possiede il clock nominale del processore, mentre il secondo un clock di frequenza molto minore. Questo perchè i segnali all'esterno dal chip sono molto più lenti rispetto ai tempi di commutazioni dei transistor presenti nell'ALU.

Per migliorare le prestazioni si sono introdotti dei miglioramenti architetturali: si è scelto di disaccoppiare il bus di sistema da quello della CPU demandando la fase di prelevamento istruzioni alla BIU (Bus Interface Unit) ed inserendo una coda di prefetch, il che consiste nel precaricare l'istruzione successiva del programma in esecuzione; è anche possibile utilizzare una memoria cache L1 apposta per il codice.

Si è inoltre scelto di aumentare il parallelismo di esecuzione introducendo una o più pipeline o utilizzando un'architettura superscalare, cioè un'architettura che permette di eseguire più istruzioni contemporaneamente.

Per quanto riguarda i sistemi multi-core il miglioramento delle prestazioni è conseguito grazie ad un più ampio utilizzo del parallelismo tra thread (strategia Thread Level Parallelism) piuttosto che ad uno tra istruzioni (Instruction Level Parallelism): quest'ultima strategia infatti ha bisogno di più complessità hardware oltre che di nuovi criteri di programmazione.

IL Modello PC

L'architettura di un PC base è formato da:

- Un processore contenente la Cpu e una floating-point unit
- Le memorie DRAM,ROM
- Un'unità di control chip (interrupt,dma)
- Una serie di dispositivi per la gestione del video, della port parallela, del floppy etc.
- Un livello di bus

L'architettura di un pc odierno però possiede fino a 3-4 livelli di bus, un microprocessore che interconnette cache L1 e un chipset (north-bridge, south-bridge) che gestisce l'interfaccia col bus del processore, gestisce la memoria DRAM, oltre a eventuali livelli di cache e ammortizza le differenze velocità di clock presenti sui vari bus gestiti .

Il chipset ha inoltre il compito di garantire la connessione con livelli di bus più bassi, ad esempio un bus per grafica veloce, o il bus di sistema (PCI) . E' presente un livello di bus ancora più basso (ISA), che può non essere presente nei portatili o in alcuni pc industriali.