

Più processori uguale più velocità?

Se un processore impiega per eseguire un programma un tempo T_1 , un sistema formato da P processori dello stesso tipo esegue lo stesso programma in un tempo $T_P = T_1 / P$?

In tal caso l'accelerazione dell'elaborazione risulterebbe essere:

$$S = T_1 / T_P = P$$

Se questa affermazione fosse vera, l'accelerazione di elaborazione possibile **sarebbe funzione lineare del numero di processori utilizzati dal sistema.**

Purtroppo questo tipo di condizione **si verifica soltanto in pochi casi molto particolari**, che corrispondono per la massima parte ad applicazioni in cui uno stesso programma deve essere eseguito molte volte con parametri iniziali diversi, allora è sufficiente che il sistema operativo distribuisca a ciascun processore i valori iniziali su cui lavorare e le CPU elaborano in maniera completamente indipendente.

Si, ma solo in qualche caso particolare

Un caso molto noto di elaborazione di questo tipo si verifica nei cosiddetti **calcoli su reticolo** nella Fisica teorica delle particelle elementari.

Un gruppo di ricercatori dell'Istituto Nazionale di Fisica Nucleare, già da molti anni, ha progettato e costruito vari prototipi di sistemi multiprocessore, molto noti nell'ambiente scientifico, in grado di eseguire con altissima efficienza calcolo parallelo nel settore specifico di loro competenza ed in qualche situazione analoga.

Questi sistemi sono ora prodotti in piccola serie e commercializzati da una ditta italiana.

Esistono, anche, particolari programmi in cui, addirittura, l'accelerazione per elaborazione parallela può crescere più che linearmente col numero dei processori usati.

Problemi dell'elaborazione parallela

Un programma contiene sempre blocchi di istruzioni tra loro indipendenti che possono essere eseguiti in parallelo e che il sistema operativo può distribuire in esecuzione a diversi processori, ma, prima o poi, i risultati parziali debbono essere messi in relazione tra loro e ciò è possibile solo a partire dal momento in cui tutti gli elementi sono disponibili.

Questa parte del programma deve necessariamente essere eseguita da un unico processore.

Problemi dell'elaborazione parallela (2)

Se l'elaborazione è affidata a più processori possono manifestarsi **due fenomeni con effetti negativi sul tempo di elaborazione.**

- Il segmento di programma affidato ad una CPU, richiede un tempo nettamente superiore agli altri, ***il carico non è uniformemente distribuito*** e tutte le altre CPU rimangono inoperose per una parte del loro tempo.
- Quando inizia la parte di programma non segmentabile in elaborazioni parallele, **i tempi di esecuzione diventano totalmente indipendenti dal numero di CPU del sistema.**

Entrambi i problemi diminuiscono drasticamente l'accelerazione reale ottenibile nell'esecuzione di un programma con l'uso di un sistema multiprocessore.

La legge di Amdahl

Una valutazione quantitativa delle prestazioni di un sistema multiprocessore può essere data utilizzando una formalizzazione dell'accelerazione che si può ottenere migliorando un sistema esistente, che **Gene Amdahl**, il progettista del famoso sistema 360 dell'IBM, propose nel 1967. La formula si riferisce all'esecuzione di un ben preciso programma.

$$S_N = \frac{T_V}{T_N} = \frac{1}{1 - f + \frac{f}{S_M}}$$

S_N è l'accelerazione del sistema migliorato T_V e T_N sono, rispettivamente, il tempo di elaborazione del "vecchio" sistema e del "nuovo", f è la frazione del tempo di elaborazione che si avvale dei miglioramenti ed S_M è l'accelerazione raggiungibile se i miglioramenti fossero efficaci su tutto il tempo di elaborazione.

Riferimenti storici

All'epoca in cui essa fu formulata, la legge di Amdahl si applicava, per esempio, al caso in cui ad un sistema si aggiungesse la cosiddetta "*unità aritmetica estesa*" che era lo "*hardware processor*" che effettuava moltiplicazioni e divisioni in parallelo. Successivamente si poteva applicare la formula per confrontare le prestazioni di processori con e senza il "*coprocessore matematico in virgola mobile*".

E' chiaro che in entrambi i casi i risultati ottenuti dipendevano in maniera critica dalla presenza nel programma delle operazioni che l'unità aggiunta velocizzava.

In tutti i confronti di prestazioni tra sistemi la struttura dei programmi ai quali si fa riferimento (*Bench-mark*) hanno un ruolo determinante

Applicazione al *multiprocessing*

La "legge di Amdahl" può essere applicata al caso del passaggio da una elaborazione mono-processore ad una multi-processore con piccoli adattamenti.

Se con f si indica la **frazione del tempo di elaborazione** che nel programma che bisogna eseguire si presta ad essere ridotto eseguendo i corrispondenti blocchi di istruzioni in parallelo e con P il **numero di processori adoperati**, l'accelerazione ottenibile S_P è data da:

$$S_P = \frac{1}{1 - f + \frac{f}{P}} = \frac{P}{P - f(P - 1)}$$

E' immediato osservare che per $f \rightarrow 1$, $S_P \rightarrow P$

La formula presuppone che si possa raggiungere un perfetto bilanciamento del carico di elaborazione sui P processori adoperati. E' ovvio che realisticamente ciò può essere vero per P abbastanza piccolo, ma è sicuramente non verificato quando P raggiunge valori elevati.

Una valutazione numerica

Se si ipotizza un programma parallelizzabile per l'80% del suo tempo di esecuzione esso viene eseguito da un multiprocessore a 16 CPU con una accelerazione di:

$$S_{16} = \frac{16}{16 - 0,80 \cdot 15} = 4$$

Se le CPU raddoppiano passando da 16 a 32, con un incremento di costo del sistema non indifferente, l'accelerazione passa solo da 4 a 4,4 e ciò, nonostante si sia fatto un esempio di programma con una bassa percentuale di blocchi di istruzioni necessariamente sequenziali.

Come si vede **in base alla formula di Amdahl, l'accelerazione di un multiprocessore cresce molto meno che linearmente col numero P dei processori**, nonostante che la relazione si fondi sull'ipotesi ottimistica di perfetto equilibrio del carico elaborativo dei processori.

Considerazioni conclusive

In pratica, salvo i casi particolari a cui abbiamo accennato all'inizio, **il limite di convenienza nell'uso di multiprocessore in applicazioni di tipo generale è limitato a poche CPU** (di solito 4) come avviene nei "server" commerciali, in cui il vantaggio deriva sostanzialmente dalla multiutenza, cioè dal fatto che il sistema di gestione, il primo, fin che è possibile, assegna ad ogni utente un processore ed, in questa situazione, il parallelismo diventa del 100%.