

Parallelismo e Haskell

Jacopo Francesco Zemella

22 agosto 2016

Prefazione

La prima domanda da porsi per comprendere al meglio le ragioni dietro il lavoro in oggetto è la seguente: come è possibile ottimizzare al meglio le risorse della macchina? La seconda è conseguentemente: esclusa la multiprogrammazione, limitata all'esecuzione concorrente dei processi, che alternative ho a disposizione, se il problema riguarda un programma solo?

Come da titolo una possibile alternativa è rappresentata dal parallelismo, una vera e propria "parallelizzazione" del lavoro mediante l'ausilio di "lavoratori" (i processori) che eseguono ciascuno una parte del lavoro (i threads) concorrentemente.

Naturalmente il passaggio dal sequenziale (ovvero programmazione priva di parallelismo) al parallelo non è decisamente indolore. Oltre a una pianificazione decisamente più complessa del programma da eseguire, subentrano numerose problematiche legate alla coordinazione dei processori o alla dimensione del problema.

Va dunque detto questa tesi non ha come finalità il rendere obsoleto il concetto di programmazione sequenziale: al contrario verranno mostrati casi in cui la parallelizzazione non solo è sconveniente, ma addirittura meno performante della sua controparte.

Lo scopo ultimo del lavoro è piuttosto mostrare in primo luogo come molti problemi, spesso anche relativamente semplici nella loro comprensione e implementazione sequenziali, risultino lenti e poco efficienti (soprattutto dal punto di vista di tempo di esecuzione) al crescere della loro dimensione. In questi casi il parallelismo trova un suo perché, arrivando a migliorare l'esecuzione ottimizzando le risorse della macchina.

Dimostrato questo concetto, subentra il problema di come implementare queste caratteristiche in un programma. Per questa seconda parte ci serviremo di un linguaggio di programmazione di comodo utilizzo, Haskell, un linguaggio funzionale avente a disposizione astrazioni e librerie che consentono di parallelizzare problemi generici già a livello di codice, a vantaggio del programmatore.

Indice

I	Calcolo Sequenziale	3
1	Algoritmi sequenziali	4
2	Macchina di Turing Deterministica	5
2.1	Complessità sequenziale	6
2.1.1	Lato pratico	7
3	Esempi	8
3.1	Ordinamento	8
3.2	Algebra Lineare	9
3.3	Grafi	9
II	Dal Sequenziale al Parallelo	10
4	Modelli Paralleli	11
4.1	Perchè parallelizzare problemi sequenziali	11
4.2	Problemi nella parallelizzazione	11
5	Efficienza negli Algoritmi Paralleli	12
6	Esempi	13
6.1	Somme Prefisse	13
6.2	Ordinamento	13
III	Programmazione Parallela	14
7	Il Parallelismo oggi	15
8	Haskell	17
8.1	Introduzione al linguaggio Haskell	17

8.1.1	Installazione	18
8.2	Perchè Haskell?	18
8.2.1	Linguaggio funzionale: un linguaggio matematicamen- te puro	19
8.2.2	Astrazione	21
8.2.3	Parallelismo in Haskell	22
9	Parallel Haskell	23
9.1	Parallelismo puro: Control.Parallel	24
9.1.1	Esempio: la serie di Fibonacci	25
9.2	Le strategie di calcolo	26
9.2.1	La monade Eval	26
9.2.2	Usare le strategie	26
9.2.3	Esempio: la funzione map in parallelo	27
9.3	Limiti di Haskell	28
IV	Implementazione in Haskell	29
10	Problemi noti	30
11	Altri esempi implementati	31

Parte I

Calcolo Sequenziale

Capitolo 1

Algoritmi sequenziali

Prima di avviarcì nella teoria del calcolo parallelo è doverosa un'introduzione nel concetto del calcolo algoritmico sequenziale.

Molti dei concetti riguardanti la parallelizzazione del codice è un'estensione della sua controparte sequenziale e figlia delle sue ovvie limitazioni nell'ambito dell'esecuzione.[1]

Solitamente quando pensiamo ad un algoritmo generico il nostro primo pensiero è quello di un programma atto ad essere eseguito su una macchina di Von Neuman, il più semplice e minimale modello di calcolatore, dotato di una singola CPU e che esegue una serie (finita) di istruzioni eseguite una dopo l'altra. Sequenzialità, dunque, significa prendere un problema e "spacchettarlo" in una sequenza di operazioni da eseguire da un singolo processore.

Poiché la nostra analisi è orientata principalmente allo studio della complessità algoritmica, ci concentreremo su un comodo modello di calcolo, specializzato in questo campo: la macchina di Turing. Esso ci permetterà di studiare gli algoritmi sequenziali e di ottenere risultati che ci torneranno utili qu

Capitolo 2

Macchina di Turing Deterministica

La macchina di Turing (MdT) è un modello ideale di calcolatore dalle meccaniche intuitive e semplici, solitamente usato più per l'analisi computazionale che per l'implementazione di calcolatori reali. Vista la sua semplicità, rappresenta uno dei modelli più utilizzati per identificare e studiare il calcolo sequenziale.

Una MdT è composta essenzialmente da:

- un'unità di controllo a stati finiti (deterministica, da cui il nome del titolo);
- un nastro diviso in celle di dimensione arbitraria contenente i dati di input/output;
- una testina capace di leggere, scrivere o cancellare dati sul suddetto nastro (inizialmente posizionata sulla cella più a sinistra del nastro).

Analogamente l'esecuzione di una MdT è suddivisa nelle seguenti fasi:

- inizialmente, trattandosi di un'automa a stati finiti, l'unità di controllo si trova in uno stato iniziale;
- viene scritta su nastro la "stringa" di input;
- a seconda dello stato corrente dell'unità di controllo e del carattere letto mediante la testina, la macchina si posiziona sullo stato successivo; da ciò dipende il comportamento della testina sulle operazioni su nastro (lettura, scrittura o nessuna operazione), comprese azioni di scorrimento sullo stesso;

- l'esecuzione termina se viene letta un'operazione non riconosciuta o se l'unità di controllo ha raggiunto uno stato finale.

La MdT ha avuto un ruolo fondamentale nell'informatica teorica: poiché è matematicamente dimostrato che per qualsiasi modello di calcolo esiste una macchina di Turing associata (tesi di Church-Turing), questo ha reso possibile definire uno standard nell'analisi computazionale degli algoritmi.

2.1 Complessità sequenziale

Quando analizziamo un algoritmo dobbiamo tenere in considerazione due caratteristiche: il fatto che risolva correttamente il problema che gli viene chiesto di risolvere (correttezza) e quante risorse impiega per essere eseguito (complessità).

La complessità di un algoritmo si basa essenzialmente su due misure esprimibili come $T_A(x)$ e $M_A(x)$, rispettivamente il tempo di calcolo e lo spazio utilizzati dall'algoritmo A su input x . In ambito informatico l'efficienza di un algoritmo è basata essenzialmente sull'utilizzo che un algoritmo fa delle risorse della macchina: il tempo è correlato alla CPU e lo spazio dalla memoria.

Algoritmi aventi la stessa finalità possono non avere il medesimo utilizzo di tempo e memoria. Per questo motivo si parla di efficienza, per distinguere al meglio la qualità di un algoritmo rispetto ad un altro.

Misurare le caratteristiche di un algoritmo non è un concetto immediato come può sembrare. L'esecuzione di un programma su un calcolatore dipende sempre da numerosi fattori: la macchina usata, il linguaggio, il compilatore, ecc. Dunque è bene definire una stima **assoluta** di quella che è la complessità di un algoritmo.

Qui entra in gioco la macchina di Turing: dal momento che ogni algoritmo può essere tradotto in modo da poter essere eseguito su una MdT possiamo sfruttarla per ottenere stime temporali come funzioni.

Tempo

In termini di tempo diciamo che un algoritmo A che opera su un generico input x restituisce un risultato dopo un tempo $T_A(x) = f(n)$, con n pari alla lunghezza dell'input x e $f(n)$ pari al numero totale di operazioni da eseguire (funzione dipendente dalle dimensioni dell'input).

In genere quando si tratta di efficienza di un algoritmo il tempo di esecuzione assume un ruolo primario rispetto alla memoria.

Memoria

In termini di spazio di memoria diciamo che un algoritmo A che opera su un generico input x necessita di un numero di "celle" di memoria pari a $M_A(x) = f(n)$, con n pari alla lunghezza dell'input x .

2.1.1 Lato pratico

Abbiamo appena derivato un metodo corretto per calcolare la complessità di un algoritmo. Sfortunatamente, in ambito reale, derivare correttamente la funzione che descriva l'utilizzo delle risorse di un algoritmo può rivelarsi un compito arduo. Oltretutto il fine ultimo del calcolo della complessità algoritmica è il confronto tra i vari algoritmi per capirne l'efficienza, dunque non è richiesta una stima esatta delle risorse utilizzate.

Per questo motivo si ricorre ad un metodo di misurazione più blando, ma comunque efficace, sfruttando la teoria del calcolo asintotico. Più precisamente ci affidiamo alla notazione $\mathcal{O}(n \log n)$ ("O grande") e a successioni di immediata comprensione per dare un'idea dell'efficienza dell'algoritmo rispetto a casi particolari di esecuzione. In particolare per ogni algoritmo riconosciamo le relative prestazioni nei seguenti casi:

- caso ottimo: caso di esecuzione che richiede all'algoritmo meno operazioni per ottenere una soluzione;
- caso peggiore: caso di esecuzione che richiede all'algoritmo il maggior numero possibile di operazioni per ottenere una soluzione;
- caso medio: caso definito statisticamente; rappresenta un valore medio di operazioni che esegue l'algoritmo rispetto ad un numero sufficientemente grande di esecuzioni.

Tra tutti i casi, ovviamente quello medio è l'elemento più utile ai fini del calcolo della complessità computazionale, poiché solitamente ci aspettiamo di eseguire un algoritmo su problemi generici. Di seguito facciamo degli esempi classici e che verranno implementati alla fine dell'elaborato.

Capitolo 3

Esempi

Di seguito verranno presentati alcuni esempi di algoritmi sequenziali, molti dei quali verranno implementati a livello di codice alla fine dell'elaborato.

3.1 Ordinamento

Quello dell'ordinamento di un insieme di oggetti confrontabili è uno degli problemi tipici del calcolo algoritmico. Ne esistono di diversi tipi, il che lo rende un esempio perfetto per confrontare le varie alternative.

Prendiamo ad esempio due algoritmi piuttosto comuni in questo campo: quicksort¹ e bubblesort². In questo caso, ad esempio i vari casi di esecuzione cambiano:

- caso migliore: $\mathcal{O}(n \log n)$ per quicksort e $\mathcal{O}(n)$ per bubblesort;
- caso peggiore: $\mathcal{O}(n^2)$ per quicksort e $\mathcal{O}(n^2)$ per bubblesort;
- caso medio: $\mathcal{O}(n \log n)$ per quicksort e $\mathcal{O}(n^2)$ per bubblesort;

Concentrandoci sul caso medio appare evidente come quicksort sia generalmente più efficiente di bubblesort.

¹In linguaggio naturale: preso un elemento x in una lista, se ne creano due: quella degli elementi minori di x e quella degli elementi maggiori. Si richiama poi l'algoritmo sulle nuove liste, dopodiché si uniscono i risultati aggiungendo x in mezzo agli stessi.

²In linguaggio naturale: si scorre la lista confrontando ogni elemento con il successivo; se il successore è minore, i due vengono scambiati di posizione. L'algoritmo viene ripetuto finché non ci sono più scambi da effettuare.

3.2 Algebra Lineare

3.3 Grafi

Parte II

Dal Sequenziale al Parallelo

Capitolo 4

Modelli Paralleli

4.1 Perchè parallelizzare problemi sequenziali

Spesso uno dei problemi principali inerenti al calcolo sequenziale è il tempo di esecuzione: si pensi al miglior algoritmo possibile su un problema semplice come l'ordinamento di un set di oggetti e si immagini di aumentare il numero di elementi in maniera spropositata. Già solo a livello teorico la

4.2 Problemi nella parallelizzazione

Capitolo 5

Efficienza negli Algoritmi Paralleli

Capitolo 6

Esempi

6.1 Somme Prefisse

6.2 Ordinamento

Parte III

Programmazione Parallela

Capitolo 7

Il Parallelismo oggi

Cominciamo a discutere della parallelizzazione in senso più concreto partendo dalla **prima legge di Moore**:

"La complessità di un microcircuito, misurata ad esempio tramite il numero di transistori per chip, raddoppia ogni 18 mesi"

Dal punto di vista industriale la creazione di processori più potenti e aventi una frequenza di clock superiore comincia a perdere significato, come dimostrato dal fatto che la potenza dei singoli core abbiano avuto un improvviso blocco verso gli inizi degli anni 2000. Questo fenomeno ha reso evidenti i limiti della legge di Moore sopracitata: di per sé i processori non sono migliorati, eppure la potenza di calcolo delle architetture non ha accennato a rallentare. Come ci spieghiamo tutto ciò?

Di certo uno sviluppo tecnologico c'è stato, si è solo spostato verso un altro campo: i multiprocessori. Raggiunti i limiti fisici legati alla capacità di elaborazione dei core, la strategia migliore (e per certi versi anche la sola disponibile) per aumentare la potenza di calcolo è stata la concentrazione di molteplici processori installati sulla stessa macchina che lavorano in parallelo.

A causa di ciò, il problema della parallelizzazione è concetto sempre più presente nel panorama informatico moderno, che può aprire svariate opportunità di ricerca anche dal punto di vista dei linguaggi di programmazione. L'idea di concedere ad un programmatore gli strumenti giusti per ottimizzare le risorse della macchina già nella compilazione del codice da eseguire è al tempo stesso affascinante e allettante. Non è quindi un caso che numerosi linguaggi si siano attrezzati per fornire strutture adatte a questo scopo.

In generale, quando si tratta di paradigmi di programmazione, distinguiamo due diverse tecniche di programmazione parallela:

- implicita: il sistema riconosce autonomamente i meccanismi da adoperare per dividere il problema in modo da poterne eseguire le parti parallelamente; di fatto il programmatore non effettua nessuna precisazione sulla natura dell'esecuzione e scrive il codice come se fosse un programma sequenziale;
- esplicita: il ruolo del programmatore è quello di partizionare il problema nel modo da lui ritenuto migliore; la stesura del codice è cruciale.

La parallelizzazione esplicita è quella che ci interessa maggiormente: vogliamo trovare un modo efficace per parallelizzare un'esecuzione in maniera diretta, definendo le parti da suddividere all'interno del problema già a livello di codice.

Capitolo 8

Haskell

Tra tutti i linguaggi che consentono una gestione delle risorse interne delle architetture multi processore, Haskell è un ottimo candidato per essere il più fruibile tra tutti quelli disponibili.

Oltre a fornire un'interessante serie di caratteristiche formali, quali la programmazione funzionale, la lazy evaluation, le funzioni higher-order e altre che vedremo in dettaglio, possiede già un ricco quantitativo di librerie e strutture che consentono la creazione di programmi paralleli già nel codice

8.1 Introduzione al linguaggio Haskell

Haskell è un linguaggio di funzionale puro caratterizzato da una tipizzazione forte e statica, creato alla fine degli anni '80, disponibile e integrabile su tutte le principali piattaforme software odierne. Possiede le caratteristiche più note dei linguaggi di programmazione (funzionali e non), tra cui la lazy evaluation, il polimorfismo e le funzioni di ordine superiore, in virtù del lambda calcolo, su cui si regge dal punto di vista matematico. Oltre a questo supporta i tipi di dati più comuni, come gli interi e i valori booleani, e permette di crearne di nuovi sotto forma di strutture dati.

Particolare di non poco conto per il nostro interesse, è già pienamente avviato nel mondo dell'elaborazione parallela e dispone di numerose librerie in open source che supportano programmazione parallela e concorrente.

Descrivere nel dettaglio un intero linguaggio di programmazione rischia di risultare troppo complicato, pertanto ci limiteremo a descriverne i punti salienti per concentrarci sul lato della programmazione parallela; un tutorial su come muovere i primi passi in Haskell è disponibile sul sito <https://www.schoolofhaskell.com/>.

8.1.1 Installazione

Per poter eseguire correttamente Haskell è necessario installare il compilatore associato, GHC, the Glasgow Haskell Compiler, disponibile online su <https://www.haskell.org/ghc/> per tutte piattaforme Linux, Windows e Mac OS.

Oltre a questo è disponibile anche il tool Cabal per l'installazione di package aggiuntivi, scaricabile dal sito <https://www.haskell.org/cabal/>.

Il package che utilizzeremo per questo lavoro saranno Parallel (<https://hackage.haskell.org/package/parallel>) per l'esecuzione parallela.

Inoltre è consigliabile un editor di programmazione adatto allo scopo. Un semplice editor di testo è sufficiente, ma è consigliabile utilizzare IDE più specifici, come Leksah o Eclipse. I file di haskell hanno tutti formato *.hs*.

8.2 Perché Haskell?

Ci sono svariati motivi legati all'utilizzo di Haskell per programmare in parallelo:

- il paradigma sfruttato da Haskell facilita la creazione di codice pulito e conciso;
- possiede una semantica relativamente semplice da acquisire;
- il codice conciso permette di evitare errori di programmazione o di rilevarne più facilmente;
- programmi molto più facili da testare
- consente un livello di astrazione molto alto
- le librerie disponibili consentono già una programmazione parallela implicita ad alto livello.

Va fatto notare che la maggior parte delle caratteristiche che sono state enunciate sono rappresentate dal semplice fatto che Haskell è un linguaggio funzionale. In generale i linguaggi funzionali hanno sempre dimostrato un ottimo utilizzo nella progettazione del software, poiché, consentono una maggiore fruibilità di altre fasi di lavorazione, quali stesura delle specifiche (test funzionali) e manutenzione del software (software inspection facilitato dal codice conciso).

8.2.1 Linguaggio funzionale: un linguaggio matematicamente puro

Andiamo più nel dettaglio per quanto riguarda le categorie dei linguaggi informatici. I linguaggi di programmazione più comuni e usati sono solitamente linguaggi *imperativi*. Si pensi ai più famosi, come C, Java e così via. Rappresentano la tecnica di programmazione più classica e "antica": ogni programma è composto da una serie di istruzioni eseguite sequenzialmente, una dopo l'altra. Le varie funzioni (in C) e metodi (in Java) che compongono i programmi sono di fatto *procedure* che lavorano in maniera iterativa.

Al contrario, Haskell è un linguaggio *funzionale*; appartiene, cioè, ad una categoria di linguaggi composta essenzialmente da funzioni matematiche pure, da cui il nome del paradigma. Una funzione in un linguaggio funzionale non esegue calcoli come nella sua controparte imperativa, ma, come in matematica, mappa elementi di un insieme (Dominio) in un altro (Codominio).

Ecco come si presenta una funzione matematica:

$$f(x) = y \tag{8.1}$$

Ad ogni elemento x appartenente al dominio X di definizione, la funzione f associa univocamente il valore y del Codominio Y . In questo caso si dice che y dipende da x in base alla relazione:

$$f : X \rightarrow Y \tag{8.2}$$

Tale concetto rappresenta la base dell'algebra e viene implementato elegantemente da tutti i linguaggi funzionali.

Osserviamo un esempio semplice in Haskell, come la funzione incremento:

```
module Esempio where
```

```
inc :: Int -> Int
inc x = x + 1
```

Il valore restituito da *inc* dipenderà dall'input (intero) in ingresso. In particolare (aperta la console ghci) otterremo:

```
Prelude Esempio> inc 2
3
Prelude Esempio> inc 4
5
Prelude Esempio> inc 100
101
```

Funzioni di ordine superiore

Ci addentriamo ora in uno dei punti chiave dei linguaggi funzionali. Come già specificato, in questo tipo di linguaggi ogni elemento che compone il programma è rappresentato da una funzione matematica. Ciò rende effettivamente potenti i linguaggi funzionali è lo sfruttamento di questa caratteristica per creare quelle che vengono definite funzioni **higher-order** o di ordine superiore.

Una funzione higher-order è un particolare tipo di funzioni che accetta altre funzioni come parametri o ne restituisce altre come risultato.

Facciamo un esempio aggiungendo al codice di prima una nuova funzione:

```
myOperator :: (Int -> Int) -> Int -> Int
myOperator f n = (f n) * 2
```

In questo caso la funzione *myOperator* accetta in ingresso una funzione da *Int* a *Int* e un valore *Int*. Il suo valore equivale al valore della funzione operata sull'intero in input moltiplicato per 2. Poiché la funzione *inc* corrisponde alle richieste, possiamo sfruttarla per fare qualche esempio di esecuzione:

```
Prelude Esempio> myOperator inc 2
6
Prelude Esempio> myOperator inc 4
10
Prelude Esempio> myOperator inc 100
202
```

Questi sono esempi basilari, ma dovrebbero rendere conto delle potenzialità inerenti ad un linguaggio funzionale come Haskell. Le funzioni di ordine superiore non solo consentono un alto livello di modularità del codice, ma aprono le porte a numerose funzionalità, tra cui la capacità di incapsulare tra di loro varie funzioni per variarne meccanismi e risultati.

Monade

Altro elemento cardine di Haskell è la **Monade**. Una monade è sostanzialmente una classe di computazioni concatenabili, e consente ad Haskell di eseguire operazioni sequenziali anche nel suo contesto funzionale.

Per rendere più chiaro il concetto, definiamo una monade come una tripla $\langle M, \text{return}, \gg= \rangle$, in cui:

- M è un costruttore di tipi;
- return è una funzione che specifica come costruire i tipi di monadi a partire dal loro contenuto;
- $\gg=$ è l'operazione di bind; presi due parametri, passa ogni valore prodotto dal primo come argomento per il secondo; è ciò che rende sequenziali le operazioni per i linguaggi funzionali.

In Haskell la classe `Monad` si presenta così¹:

```
class Monad m where
  (>>=) :: m a    -> (a  -> m b) -> m b
  (>>)   :: m a    -> m b -> m b
  -- (>>) è come (>>=) ma scarta il risultato dell'operazione
  return :: a      -> m a
  fail    :: String -> m a
```

8.2.2 Astrazione

Il fatto che ogni elemento di Haskell sia una funzione matematica che può prenderne altre in ingresso o restituirne altre in uscita consente a una funzione di non avere una definizione esplicita di come funziona esattamente una funzione per poter essere sfruttata. Ogni elaborazione può essere sfruttata e intercambiata a piacimento e viene decisa solo in fase di esecuzione.

Tale caratteristica prende il nome di astrazione, la capacità di creare elementi

¹In Haskell le lettere minuscole nella definizione dei tipi indicano sempre tipi generici; una funzione `f :: a -> b` indica una relazione tra un tipo generico `a` e un'altro tipo generico `b`, non necessariamente della stessa natura.

il cui funzionamento interno è definito in maniera implicita e non diretta. L'astrazione è consente di modulare il codice a vantaggio della progettazione generale e permette la creazione di strutture ad alto livello intercambiabili e di semplice utilizzo. Un buon livello di astrazione è generalmente indice di una buona programmazione.

8.2.3 Parallelismo in Haskell

All'inizio del capitolo abbiamo distinto due categorie di parallelismo, implicito ed esplicito, stabilendo di voler concentrare la nostra attenzione sulla seconda. Nelle librerie che utilizzeremo Haskell garantisce una sorta di via di mezzo, una forma definita semi-esplicita: per l'esecuzione parallela sono richiesti dall'utente solo alcuni aspetti della coordinazione tra le varie partizioni del programma. Tuttavia ciò non è assolutamente un fatto negativo: cedendo al sistema la maggior parte del lavoro "sporco", ci consentirà di focalizzarci solo sulla stesura del codice.

Il nostro lavoro sarà solo ed esclusivamente partizionare il problema all'interno del codice in sezioni che possano essere eseguite parallelamente, senza precisazioni sulle caratteristiche della macchina che andrà ad eseguire il codice, nemmeno il numero di processori di cui dispone.

Ovviamente l'obiettivo è di mantenere i processori in attività sul problema limitando le interazioni tra gli stessi. Ovviamente i problemi derivanti da questa programmazione non mancano. In primis vi è la dipendenza che può intercorrere tra i dati: due sezioni possono condividere una stessa sezione di memoria o essere l'una in attesa di un risultato offerto dall'altra. Ciò porta ad una sorta di ritorno alla sequenzialità non accettabile visto il nostro obiettivo. In secondo luogo vi è il concetto di granularità dei threads, ovvero la dimensione delle partizioni che andremo a creare. Threads troppo piccoli rischiano di diventare un collo di bottiglia per l'esecuzione, che impegna i processori principalmente a coordinarsi piuttosto che a eseguire lavoro utile, vanificando (o anche rendendo obsoleti) gli sforzi della parallelizzazione.

Capitolo 9

Parallel Haskell

Abbiamo affermato che la finalità di sfruttare Haskell in parallelo (o come viene comunemente chiamato, Parallel Haskell), è di suddividere il lavoro e di eseguirlo su più processori contemporaneamente.

L'esecuzione in parallelo non è implicita nel codice e deve quindi passare con le corrette impostazioni prima a livello di compilatore, e successivamente a livello di esecuzione. Bisogna dunque assicurarsi di aver scaricato l'ultima versione di GHC, o di averlo aggiornato di conseguenza, e, una volta terminato il codice eseguire i seguenti passi:

- compilare il codice con l'opzione `-threaded`; esempio `ghc -threaded foo.hs -o foo`;
- una volta ottenuto il compilato, va mandato in esecuzione con l'opzione `+RTS -N4`; nello specifico `RTS` lega l'esecuzione con un sistema Real-Time (per la gestione autonoma di threads, memoria e così via), mentre l'opzione `-N4` specifica il numero di core da utilizzare, in questo caso quattro (omettere il numero di core da usare consente all'esecuzione di utilizzare tutti quelli presenti nella macchina); esempio `foo +RTS -N4`;
- per analizzare anche le statistiche temporali dell'esecuzione, così da confrontare la parallelizzazione con la controparte sequenziale, è anche utile compilare il codice aggiungendo l'opzione `-rtsopts` ed eseguirlo aggiungendo l'opzione `-sstderr`; l'esecuzione produrrà un file *stderr* contenente le specifiche sul tempo di esecuzione e sulle risorse impiegate.

9.1 Parallelismo puro: Control.Parallel

Cominciamo a descrivere la parallelizzazione esplicita in Haskell partendo dalla libreria più importante in questo ambito, ciò che ci consentirà di trasformare il nostro codice sequenziale in codice parallelo.

La libreria `Control.Parallel` si compone essenzialmente di due funzioni, `par` e `pseq`, così definite:

```
par :: a -> b -> b
pseq :: a -> b -> b
```

La funzione `par`

La funzione `par f1 f2`, esprimibile anche come `f1 'par' f2`, si traduce sostanzialmente in: "l'esecuzione di `f1` può essere parallelizzata con `f2`; il valore di ritorno è il risultato di `f2`".

Questa funzione viene usata per effettuare l'esecuzione parallela di due elementi, di cui il valore del primo non è richiesto immediatamente.

La funzione `pseq`

La funzione `pseq f1 f2`, esprimibile anche come `f1 'pseq' f2`, è la controparte parallela della funzione `seq`, della libreria standard di Haskell. Nel caso di `seq` si tratta di una funzione sfruttata per ottimizzare le prestazioni dell'esecuzione, mediante la gestione della lazy evaluation: `seq f1 f2` restituisce il valore di `f2` solo dopo che entrambe le funzioni in input hanno terminato il loro ciclo di esecuzione¹.

Analogamente la funzione `pseq f1 f2` si regge sullo stesso principio di base, ma con una particolarità in più: in genere, quando si tratta di parallelismo, vogliamo sfruttare il controllo sulla lazy evaluation in modo da valutare `f1` prima di `f2`, magari perché abbiamo già specificato che i due vanno eseguiti in parallelo.

In tal caso `seq` non è sufficiente, poiché legata a entrambi gli argomenti. Con `pseq`, invece, ci assicuriamo che l'esecuzione avvenga nell'ordine corretto.

¹Piccola nota: ciò non significa che `f2` viene eseguita dopo `f1`; si tratta solo di un modo per impedire inutili valutazioni pigre

9.1.1 Esempio: la serie di Fibonacci

Proviamo con l'"Hello World" degli algoritmi, uno semplici da implementare. Ricavare l' n -esimo elemento della serie di Fibonacci in Haskell diventa:

```
fib :: Int -> Int

fib 0 = 1
fib 1 = 1
fib n = (fib (n-1)) + (fib (n-2))
```

Parallelizziamo ora il codice usando pseq e par:

```
fibpar :: Int -> Int
fibpar 0 = 1
fibpar 1 = 1
fibpar n = (n1 'par' n2) 'pseq' (n1 + n2)
           where
             n1 = fibpar (n-1)
             n2 = fibpar (n-2)
```

In questo caso, dato un generico intero n , viene parallelizzata l'esecuzione di Fibonacci su $n-1$ insieme a quella di $n-2$, specificando che la loro somma deve attendere che restituiscano un risultato.

Notiamo che di per sé gli algoritmi sono concettualmente identici: effettuano la ricorsione due volte per chiamata e uniscono i risultati una volta terminate entrambe le esecuzioni.

Confrontiamone ora il tempo di esecuzione effettuato su input 46: il codice sequenziale ha restituito il risultato dopo 260 secondi, quello parallelo in 58. Ricaviamo quindi uno speedup di:

$$Speedup = \frac{260sec}{58sec} = 4.48 \simeq 450\% \quad (9.1)$$

9.2 Le strategie di calcolo

Un altro modo per parallelizzare l'esecuzione in Haskell è l'utilizzo delle **strategie**. Una strategia è essenzialmente una funzione in una particolare monade², `Eval`, che prende in ingresso il tipo da associare alla stessa. Il meccanismo in Haskell è il seguente:

```
type Strategy a = a -> Eval a
```

9.2.1 La monade `Eval`

La monade `Eval` è una struttura contenuta nel package `Control.Parallel.Strategies`. Vediamo nel dettaglio come la sua composizione:

```
data Eval a
instance Monad Eval
```

```
rpar :: a -> Eval a
rseq :: a -> Eval a
```

```
runEval :: Eval a -> a
```

Si compone principalmente di due funzioni, **rpar** e **rseq**, semanticamente e funzionalmente identiche alle sopracitate **par** e **pseq** del package `Control.Parallel` (si veda il capitolo precedente). L'unica differenza sostanziale è il tipo che viene restituito: trattandosi di funzioni in seno a una monade, restituiranno, per l'appunto, un elemento `Eval` (con relativo tipo).

Vi è poi un'altra funzione, **runEval**, finalizzata ad effettuare la computazione della monade e restituirne il risultato.

9.2.2 Usare le strategie

Avviamoci ora verso l'utilizzo del tipo **Strategy**. Usare una strategia significa prendere in ingresso una struttura dati, utilizzare le funzioni **rpar** e **rseq** per creare parallelismo e restituire il valore del parametro in ingresso.

In questo caso le funzioni di `Eval` possono essere riscritte in funzione delle strategie:

```
rpar :: Strategy a
rseq :: Strategy a
```

²Si veda il capitolo "Perchè Haskell per un chiarimento sulle monadi."

Chiariamo il loro utilizzo con un esempio. Immaginiamo di aver creato una funzione f e di aver poi deciso di eseguirla sfruttando la strategia s . Ricavare il risultato di f diventa:

```
runEval (s f)
```

Spesso questa scrittura viene sostituita da una funzione più compatta, *using*:

```
using :: Strategy a -> a
using s x = runEval (s x)
```

Possiamo quindi riscrivere la funzione precedente come:

```
s 'using' f
```

Un corretto utilizzo delle strategie non è solo un metodo stringente per definire parallelismo. Grazie alla potenza delle astrazioni di Haskell è possibile parallelizzare un algoritmo già esistente semplicemente applicandovi una strategia per costruire il risultato.

9.2.3 Esempio: la funzione map in parallelo

Vediamo ora come l'utilizzo delle strategie possa rendere la parallelizzazione concisa ed efficace. La funzione `map` prende in ingresso due parametri: una funzione ($f :: a \rightarrow b$) e una lista di elementi di tipo a . Il nostro intento è di rendere parallela questa operazione senza modificare l'algoritmo di partenza.

La funzione più semplice finalizzata a questo intento è:

```
parMapList :: (a -> b) -> [a] -> [b]
parMapList f ls = map f s 'using' parList rseq
```

Osserviamo nel dettaglio le componenti della funzione:

- `map f s` è l'algoritmo a cui applicare la strategia
- `parList rseq` è la strategia da sfruttare; `rseq` l'abbiamo già menzionata, `parList`, invece, è una strategia già implementata in `Control.Parallel.Strategies`, che presa una strategia su valori a ne restituisce un'altra sulle liste di elementi di tipo a ; in questo caso prende la strategia `rseq` e ne restituisce una che applica la suddetta ad ogni elemento di una lista

Osserviamo alcuni elementi fondamentali: prima di tutto non abbiamo creato nessuna funzione *ex novo*, abbiamo solo assemblato algoritmi preesistenti e vi abbiamo applicato un meccanismo di calcolo; in secondo luogo abbiamo messo in rilievo che è possibile creare strategie di livello superiore senza particolari assunzioni su tipi e strutture.

9.3 Limiti di Haskell

Come è stato affermato in questo capitolo, non esiste il linguaggio perfetto, e anche Haskell non è immune a questa regola. Pur essendo comodo e di facile utilizzo manca di una gestione diretta della memoria, almeno a livello virtuale, come invece consente l'utilizzo di C, grazie ai suoi puntatori. Anche per questo motivo è bene dosare bene i linguaggi da utilizzare: non è difficile immaginare che programmi relativamente semplici, che si poggiano su algoritmi minimali, trovino una migliore applicazione in un linguaggio imperativo come C piuttosto che in Haskell.

Analogamente un programma in Parallel Haskell eseguito su un problema di piccole dimensioni potrebbe creare colli di bottiglia tra overhead e comunicazione tra processori di gran lunga superiori all'effettiva esecuzione del problema rispetto alla sua controparte sequenziale (si ricordi il concetto di granularità).

Tuttavia, tolti i problemi legati ad una cattiva progettazione, dare priorità all'ottimalità dell'esecuzione non sempre è la scelta migliore. Da questo punto di vista il linguaggio macchina sarebbe la scelta migliore per programmare, eppure è paradossale pensare che si programmi anche solo una semplice sommatoria in un linguaggio così a basso livello.

Come per molte cose, quindi, un buon compromesso è sempre la scelta migliore. Haskell non è stato scelto per essere il più ottimale dei linguaggi, ma perché consente al programmatore una gamma vastissima di scelte progettuali sposate a una buona implementazione della parallelizzazione esplicita.

Parte IV

Implementazione in Haskell

Capitolo 10

Problemi noti

Capitolo 11

Altri esempi implementati

Bibliografia

- [1] Io me medesimo. “Lor Anittal”. In: *Boh 1* (1 2016), pp. 2–3.