

Scalabilità oltre il desktop con le applicazioni parallele MATLAB

A cura di **A. Chakravarti**, MathWorks, Inc. - **P. S. Ivanov** e **J. Rorison**, University of Bristol

Introduzione

Le tecnologie di calcolo parallelo rendono disponibili i mezzi necessari per velocizzare la soluzione dei problemi di calcolo tramite risorse hardware quali workstation multicore, cluster, grid e cloud. Lo sfruttamento di risorse hardware aggiuntive consente di risolvere problemi estremamente complessi e può aiutare a ottenere notevoli vantaggi competitivi e in termini di ricerca.

Per gli ingegneri che hanno esigenze di maggiore velocità di esecuzione per i loro programmi informatici o che devono elaborare set di dati di grandi dimensioni, sarebbe ottimale se ogni programma usasse automaticamente il parallelismo. Tuttavia, la parallelizzazione automatica è una funzionalità che attualmente non esiste ed è oggetto della ricerca informatica di base. Pertanto, il compito di utilizzare la parallelizzazione per eseguire programmi su più core è condiviso tra i progettisti di linguaggi/ambienti di programmazione e i loro utenti.

Questo articolo esplora le aspettative che gli ingegneri nutrono nei confronti delle tecnologie di calcolo parallelo e descrive come i progettisti di MATLAB hanno cercato di soddisfarle.

Requisiti relativi alle capacità di calcolo parallelo

Il compito principale degli ingegneri è rappresentato dalla risoluzione di problemi complessi nell'ambito dei domini tecnici di loro competenza. Sebbene alcuni di loro possano essere anche programmatori esperti, la maggior parte preferirebbe essere al riparo dalle sottigliezze della programmazione parallela, quali il multithreading, la sincronizzazione e la gestione dei dati tra cluster. Per questo motivo un ambiente di calcolo parallelo deve facilitare il più possibile la scrittura e l'utilizzo di programmi paralleli.

Gli ingegneri utilizzano una varietà di sistemi operativi e piattaforme hardware e la loro esigenza è di non dover modificare il codice quando eseguono la migrazione di applicazioni da un sistema operativo all'altro o da un sistema desktop multicore a un cloud di grandi dimensioni. Analogamente, anche i loro colleghi e clienti non dovrebbero mai essere costretti ad apportare modifiche al codice. La scalabilità e la portabilità sono requisiti chiave per un ambiente di calcolo parallelo, in quanto la maggior parte degli ingegneri auspica che le proprie applicazioni possano utilizzare le risorse disponibili senza problemi.

Un grosso impedimento per qualsiasi ingegnere che desideri utilizzare hardware remo-

to in cluster è rappresentato dalla necessità di avere conoscenze specifiche relative al cluster. La maggior parte degli ingegneri preferirebbe che fosse l'amministratore del cluster a scrivere gli script specifici, impostare le variabili di ambiente e gestire le code dei job. La separazione tra le attività dell'utente e quelle dell'amministratore è dunque un requisito importante.

Scalabilità sul desktop e oltre: le sfide delle tecnologie specializzate

Le tecnologie di calcolo parallelo attualmente disponibili sono diverse. Alcune, come ad esempio Intel[®] TBB e Cilk, consentono ai programmatori di scrivere programmi paralleli che utilizzano computer multicore. Tuttavia quegli stessi programmi non sono scalabili al punto di poter utilizzare risorse remote come i cluster. Spesso è necessario riscriverli affinché utilizzino altre tecnologie quali l'MPI (Message Passing Interface), che sono complesse e richiedono competenze specialistiche. Questo flusso di lavoro non soddisfa un requisito importante, ossia la scalabilità dei programmi paralleli da workstation a cluster senza che sia necessario riscrivere il codice.

Le tecnologie specializzate come l'MPI presentano l'inconveniente ulteriore di richiedere all'utente di un programma parallelo di avere una qualche conoscenza del sistema su cui verrà eseguito. Di conseguenza la portabilità del codice risulta ridotta, come pure il numero di persone che possono utilizzarlo.

L'approccio di MATLAB[®] al calcolo parallelo scalabile

Il linguaggio di programmazione MATLAB è separato dall'ambiente di esecuzione: ciò significa che lo stesso programma parallelo può essere eseguito sia su workstation multicore che su cluster, grid e cloud. Un ingegnere che scrive un programma parallelo in MATLAB non è tenuto a conoscere l'ambiente in cui verrà di fatto eseguito il programma. MATLAB offre diversi livelli di controllo al programmatore che desidera convertire in programma affinché sia eseguito efficacemente in parallelo. Per alcuni programmi non è necessario riscrivere il codice, mentre per altri occorre usare API di basso livello. Le tecniche di programmazione più comunemente usate prevedono l'aggiunta di annotazioni al codice. Ad esempio, un loop per con iterazioni indipendenti può essere annotato come loop parfor. Durante l'esecuzione, l'ambiente di calcolo tenterà di eseguire le

iterazioni del loop in parallelo su più worker MATLAB (un worker è un motore di esecuzione eseguito in background su una workstation o un cluster).

La scalabilità di un programma parallelo MATLAB da workstation a cluster non richiede conoscenze sul cluster in quanto MATLAB consente di separare i ruoli di utente e amministratore del cluster. L'amministratore salva le informazioni relative al cluster in un file di configurazione (ad esempio, come inoltrare job e trasferire dati) e lo invia a tutti gli utenti del cluster. Di fatto, un utente potrebbe ricevere diverse configurazioni, una per ciascuna risorsa remota. L'utente importa le configurazioni nell'interfaccia utente di MATLAB e ne seleziona una come risorsa su cui eseguire il programma parallelo MATLAB.

Il tipico flusso di lavoro utilizzato da un ingegnere che desideri risolvere un problema tecnico molto complesso in MATLAB è il seguente: l'utente scrive un programma seriale e quindi lo parallelizza utilizzando costrutti quali parfor. L'utente sottopone il programma a test e debug sulla propria workstation utilizzando input di dimensioni ridotte. Infine l'utente aumenta le dimensioni degli input inseriti nel programma, importa una configurazione per un cluster remoto ed esegue di nuovo il programma su quel cluster.

La portabilità dei programmi MATLAB su diversi sistemi operativi e piattaforme hardware facilita la condivisione e la distribuzione dei programmi paralleli. Ad esempio, un programma sviluppato su una workstation Windows può essere eseguito su un cluster Linux o condiviso con colleghi che lavorano su laptop Macintosh.

Una storia vera: programmi MATLAB paralleli su grid EGEE

Il gruppo di Optoelettronica dell'Università di Bristol è impegnato in una ricerca sui VCSEL (Vertical-Cavity Surface-Emitting Lasers, laser a cavità verticale che emettono da una superficie), ampiamente utilizzati nelle reti per telecomunicazioni su fibra ottica. Il gruppo ha sviluppato una nuova generazione di VCSEL con cristalli fotonici (PC-VCSEL). Per eseguire simulazioni numeriche utilizzando modelli di PC-VCSEL, sono stati utilizzati i solutori MATLAB per le equazioni differenziali scalari di Helmholtz alle derivate parziali in due dimensioni e per le normali laser rate equation differenziali.

Il tempo approssimativo per la soluzione delle equazioni dei modelli era compreso tra i 10 e i 700 minuti per alcuni modelli e tra le 4 e le 60 ore per altri. Poiché questi modelli

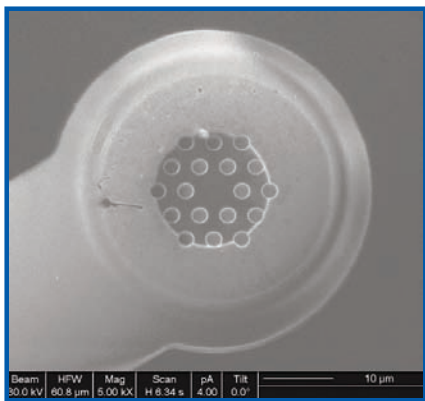
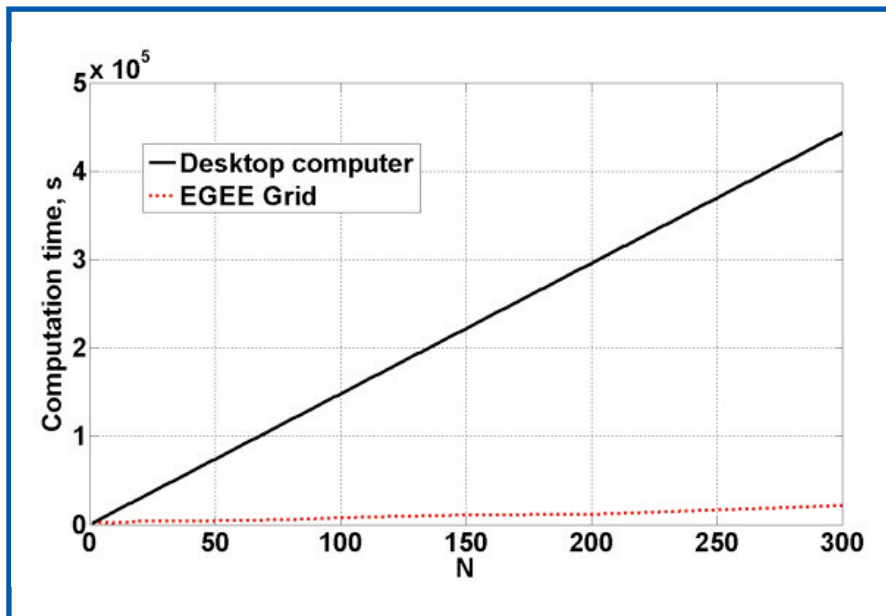


Fig. 1 - Vista dall'alto di un PC-VCSEL. Il cristallo fotonico presenta fori di $2 \mu\text{m}$ di diametro e spaziatura di $4 \mu\text{m}$ da centro a centro.

Fig. 2 - Confronto tra il tempo di calcolo sul core di una singola workstation e sul grid EGEE.



utilizzavano molti parametri di input, il calcolo delle caratteristiche del PC-VCSEL e la loro ottimizzazione richiedeva la soluzione di centinaia di equazioni. Il problema era evidentemente troppo vasto per un sistema desktop, in quanto l'esecuzione dei calcoli su una workstation del laboratorio avrebbe richiesto diversi giorni.

I ricercatori hanno parallelizzato il programma scritto in MATLAB strutturandolo come un job che calcolava i parametri dei modi ottici dei PC-VCSEL N volte. Di conseguenza si avevano N task, ciascuno dei quali calcola-

va i parametri dei modi ottici. I ricercatori hanno prima sottoposto il programma a test e debug utilizzando più MATLAB worker su una workstation. Una volta stabilita la correttezza del programma MATLAB parallelizzato, hanno deciso di eseguirlo su un sistema grid fornito da EGEE (Enabling Grids for E-science), il consorzio che mette oltre 70000 core di processori a disposizione di utenti in tutto il mondo. Utilizzando una porzione di questa infrastruttura, il tempo di calcolo di 300 task si riduceva da oltre 5 giorni a solo 6 ore: una velocità di 21 volte superiore.

Conclusioni

Gli ambienti di programmazione parallela che consentono di risolvere problemi tecnici su computer desktop e oltre sono un fattore fondamentale per lo svolgimento di attività tecniche. MATLAB fornisce la capacità di parallelizzare facilmente i programmi e quindi di andare oltre i sistemi desktop senza modificare i programmi o scrivere codice supplementare.

Per ulteriori informazioni:
www.mathworks.com

SOFTWARE AD ALTE PRESTAZIONI PER LE NUOVE TECNOLOGIE.
NON POTRETE PIÙ FARNE A MENO!

Grazie ai nostri prodotti di simulazione CAE (Computer Aided Engineering) centinaia di aziende italiane hanno ridotto i loro tempi di progettazione ottenendo notevoli vantaggi economici.

La soddisfazione dei nostri clienti deriva anche dall'accuratezza e facilità d'uso dei nostri software e dal tempismo ed efficacia del nostro supporto tecnico on-line.

Richiedi informazioni su come SmartCAE può risolvere anche i tuoi problemi di ingegneria.

Analisi Statica Lineare e Non-lineare + Analisi Termica + Dinamica + Vibrazioni + Shock + Materiali Compositi + Simulazione di Impatto + Fatica + Cinetodinamica + Fluido Dinamica Computazionale.

Contattaci per ottenere una prova gratuita nella tua Azienda chiamando lo **0574-404642** o inviando una e-mail a info@smartcae.com

 **SmartCAE**
 Simulate more, innovate faster

SmartCAE srl
www.smartcae.com
info@smartcae.com