

BOINC

Una piattaforma per il calcolo distribuito

Matteo Cicuttin

3 Giugno 2010

Agenda

Cosa tratteremo:

- ▶ Parallelismo: dove lo trovo e come lo sfrutto?
- ▶ Overview di BOINC
- ▶ Alcuni dettagli sull'infrastruttura
- ▶ BOINC vs. GRID
- ▶ Cenni minimali alle API
- ▶ Conclusioni

Parallelismo: dove
lo trovo e come lo
sfrutto?

Overview di
BOINC

Alcuni dettagli
sull'infrastruttura

BOINC vs. Grid

Conclusioni

Dove possiamo trovare il parallelismo?

Il parallelismo ha molte faccie, caratterizzate dalla complessità delle operazioni che eseguiamo parallelamente.

In modo grossolano possiamo identificare il parallelismo:

- ▶ come “intrinseco” (avete un termine migliore?)
- ▶ a livello di “sottoproblemi”
- ▶ a livello “di istanze”

Cosa significa?

Parallelismo “intrinseco”

Il parallelismo “intrinseco” lo cerchiamo a livello di somme, moltiplicazioni...

Esempio: l'algoritmo Velocity Verlet, per integrare le equazioni del moto.

- ▶ Passo 1: $x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)(\Delta t)^2$
- ▶ Passo 2: $v(t + \frac{\Delta t}{2}) = v(t) + \frac{1}{2}a(t)(\Delta t)$
- ▶ Passo 3: Ricavo $a(t + \Delta t)$
- ▶ Passo 4: $v(t + \Delta t) = v(t + \frac{\Delta t}{2}) + \frac{1}{2}a(t + \Delta t)(\Delta t)$

E allora?

Parallelismo “intrinseco”

Tipicamente non abbiamo a che fare con moto unidimensionale...

- ▶ Passo 1: $\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)(\Delta t)^2$
- ▶ Passo 2: $\vec{v}(t + \frac{\Delta t}{2}) = \vec{v}(t) + \frac{1}{2}\vec{a}(t)(\Delta t)$
- ▶ Passo 3: Ricavo $\vec{a}(t + \Delta t)$
- ▶ Passo 4: $\vec{v}(t + \Delta t) = \vec{v}(t + \frac{\Delta t}{2}) + \frac{1}{2}\vec{a}(t + \Delta t)(\Delta t)$

Ci piacerebbe fare le operazioni tra vettori in parallelo, e non componente a componente.

Strumento per sfruttarlo: le unità vettoriali che ci sono in ormai tutte le CPU, ad esempio AltiVec (PowerPC), VIS (UltraSPARC), MMX/SSE/(AVX) (Intel/AMD) oppure, un po' più ad alto livello, GrandCentral.

Parallelismo: dove lo trovo e come lo sfrutto?

Overview di BOINC

Alcuni dettagli sull'infrastruttura

BOINC vs. Grid

Conclusioni

Parallelismo a livello di sottoproblemi

Calcolo distribuito
con BOINC

Matteo Cicuttin

Parallelismo: dove
lo trovo e come lo
sfrutto?

Overview di
BOINC

Alcuni dettagli
sull'infrastruttura

BOINC vs. Grid

Conclusioni

Idea: scomporre un problema in più sottoproblemi.

Abbiamo visto parecchi esempi durante il corso ed in laboratorio, ad esempio la risoluzione parallela dell'equazione di Poisson.

Strumento per sfruttarlo: Message Passing.

Curiosità: Tra non molto forse vedremo i meccanismi di message passing implementati in hardware all'interno delle CPU, andate a vedere questo prototipo di Intel:

<http://techresearch.intel.com/articles/Tera-Scale/1826.htm>

Parallelismo a livello di istanze

Spesso è necessario risolvere uno stesso problema partendo da un certo numero di differenti condizioni iniziali, ad esempio vogliamo vedere come reagisce un organismo a diverse molecole (farmaci). Possiamo allora far girare in parallelo tutte le differenti istanze del problema.

Oppure, abbiamo enormi moli di dati da elaborare: affidiamo l'elaborazione di sottoinsiemi del dataset a molti computer differenti (Seti@HOME, che tanto per cambiare fa pesante uso della FFT).

Strumento per sfruttarlo: GRID / Calcolo distribuito.

BOINC = Berkeley Open Infrastructure for Network Computing

BOINC entra in gioco quando si parla di “parallelismo a livello di istanze”. Ma cos'è? L'idea di fondo è la seguente: ogni macchina collegata si scarica del lavoro da fare (le *workunits*) ed una volta svolto restituisce i dati elaborati.

- ▶ Nasce a Berkeley nell'ambito di Seti@HOME
- ▶ Si è visto che si adattava molto bene anche ad altri progetti
- ▶ È nata così una piattaforma di calcolo distribuito di uso generale

L'infrastruttura

L'intera infrastruttura di BOINC ruota attorno al “progetto”, ed è caratterizzata dai seguenti componenti:

- ▶ Il partecipante
- ▶ Il frontend
- ▶ Il backend

Vediamo meglio il tutto con l'aiuto di un'immagine

BOINC vs. Grid computing

Calcolo distribuito
con BOINC

Matteo Cicuttin

Parallelismo: dove
lo trovo e come lo
sfrutto?

Overview di
BOINC

Alcuni dettagli
sull'infrastruttura

BOINC vs. Grid

Conclusioni

BOINC è una piattaforma di computing volontario: ognuno condivide quante risorse gli pare e per quanto tempo vuole.

Questo, assieme ad altri fattori, lo distingue in modo abbastanza netto dal Grid computing.

Vediamo quali sono alcune delle problematiche per capire le differenze tra le due cose.

BOINC vs. Grid: Connettività (1)

- ▶ BOINC: ogni “unità di calcolo” è connessa alla rete con connessioni dial-up o ADSL
- ▶ Grid: generalmente tutti i sistemi sono interconnessi da link dedicati ad alta velocità

Quindi? BOINC deve tenere conto esplicitamente la “volatilità” delle connessioni.

BOINC vs. Grid: Connettività (2)

- ▶ BOINC: i computer dei vari utenti possono essere dietro NAT, firewall e schifezze di ogni genere
- ▶ Grid: le interconnessioni tra i vari sistemi che compongono il grid sono “studiate” e pianificate

Quindi? L'infrastruttura di BOINC, per ridurre al minimo i problemi, è costruita secondo un modello “pull”.

BOINC vs. Grid: Problematiche amministrative

Calcolo distribuito
con BOINC

Matteo Cicuttin

Parallelismo: dove
lo trovo e come lo
sfrutto?

Overview di
BOINC

Alcuni dettagli
sull'infrastruttura

BOINC vs. Grid

Conclusioni

- ▶ BOINC: gira sui computer di utenti qualsiasi, che in genere non ne sanno nulla di HPC
- ▶ Grid: i sistemi componenti il grid sono grossi e complessi, *generalmente* sotto il dominio amministrativo di staff con buone competenze di HPC

Quindi? Affinché il sistema possa avere successo, almeno i client devono essere operativi nel giro di pochi click

BOINC vs. Grid: Utenti furbi

- ▶ BOINC: gira sui computer di utenti qualsiasi, quindi potrebbe essere che qualcuno a cui non piace un dato progetto si metta a mandare indietro risultati falsi
- ▶ Grid: si sa in modo chiaro chi parla con chi

Quindi? BOINC deve supportare un meccanismo di verifica dei risultati ricevuti

BOINC vs. Grid: ...

Calcolo distribuito
con BOINC

Matteo Cicuttin

Parallelismo: dove
lo trovo e come lo
sfrutto?

Overview di
BOINC

Alcuni dettagli
sull'infrastruttura

BOINC vs. Grid

Conclusioni

Abbiamo capito alcune delle problematiche, ma ovviamente non sono esaurite qui (computer che si accendono e si spengono, sicurezza generale del sistema, ...)
E le problematiche di carattere non tecnico? Gli utenti non arrivano senza pubblicizzare il progetto...

Una slide sulle API

Ovviamente ci deve essere il modo per interfacciarsi con un'infrastruttura simile, dunque come si fa? C'è un'API apposita.

- ▶ Inizializzazione
- ▶ I/O Wrappers
- ▶ Checkpointing
- ▶ Richiesta di attenzione da parte dell'utente
- ▶ Grafica
- ▶ Credit reporting
- ▶ ...

BOINC:

- ▶ non è "la soluzione definitiva", ma uno strumento a supporto del calcolo distribuito
- ▶ non si usa indipendentemente dalle altre tecnologie, ma assieme a loro
- ▶ è un'infrastruttura particolarmente adatta dove non ci sono le risorse economiche per acquisire un supercalcolatore.