



Prima domanda:

Che differenza c'è tra:

un **Calcolatore Parallelo** e un **Sistema per il Calcolo Distribuito**



Un calcolatore parallelo e'...



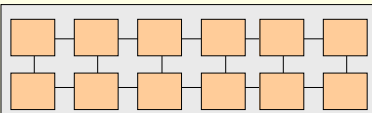
IBM Roadrunner
12960 processori IBM Cell

Performance sostenuta: **1.105** Pflops
Picco di performance: 1.456 Pflops


Los Alamos Nat. Lab

Calcolatore Parallelo

1 Pflops = 10^{15} flop/sec



Un calcolatore parallelo e'...

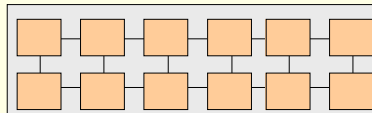


Jaguar
37538 processori AMD 4core

Performance sostenuta: **1.059** Pflops
Picco di performance: 1.381 Pflops

Oak Ridge Nat. Lab.

Calcolatore Parallelo



Top five

	Cost.	computer	sito	Nprocs	Rete conn	Linpack bench. (Pflops)
1	IBM	Roadrunner	LANL (USA)	129600 core PowerXCell 3.2 GHz	Infiniband	1.105
2	CRAY	Jaguar	ORNL (USA)	150152 core AMD Opteron 4core 2.3 GHz	XT4 (custom)	1.059
3	SGI	Pleiades	NASA	51200 core Intel Xeon 4 core 3.0 GHz	Infiniband	0.487
4	IBM	Blue Gene/L	LLNL (USA)	212992 core Power PC 0.7 GHz	proprietary	0.478
5	IBM	Blu Gene/P	ANL (USA)	163840 core Power PC 0.85 GHz	proprietary	0.450

5

Alcune statistiche dalla Top500

Architettura:

- 410 cluster
- 88 MPP
- 2 altri

Processori:

- 369 Intel Xeon
- 60 AMD Opteron
- 60 Power PC
- 9 Intel Itanium
- 2 altri

Rete:

- 282 Gbit Ethernet
- 141 Infiniband
- 42 proprietary
- 10 SP Switch
- 10 Myrinet
- 15 altri

Num. Proc:

- 290 2049-4096
- 96 4097-8192
- 61 1025-2048
- 28 8193-16000
- 21 > 16000
- 4 < 1024

6

Un calcolatore parallelo e':

Un sistema di unità processanti **omogenee**

strettamente collegate

che **comunicano**

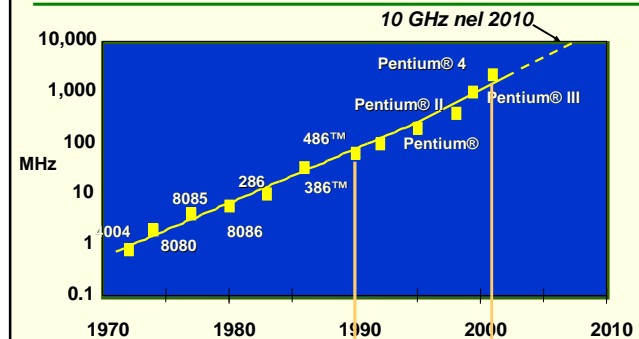
Risorse dedicate e centralizzate

per risolvere problemi su larga scala

in maniera efficiente

7

Legge di Moore



La velocità dei processori raddoppia ogni 18 mesi (crescita di 128 volte in 10.5 anni)

8

E il bandwidth delle reti?

Un esempio: La rete GARR e' la rete che connette tutte le istituzioni scientifiche italiane (universita', centri CNR, osservatori, ENEA, INFN)



Evoluzione del bandwidth

GARR	1994	2Mbit/sec
GARR-2	1998	32Mbit/sec
GARR-B	2002	2.5Gbit/sec
GARR-G	2006	10Gbit/sec

Crescita di
 1000 volte in 8 anni
 (x2 in 9 mesi)

9

IDEA



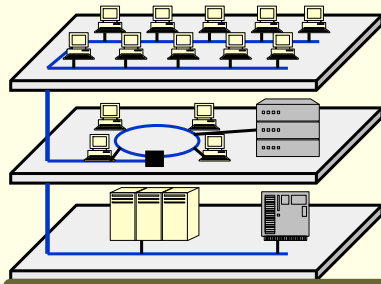
fattore di **crescita** della
velocita' delle reti di comunicazione
molto maggiore
 del fattore di **crescita** della
velocita' dei processori



Posso considerare tutti i **calcolatori**
connessi ad una rete come una **unica**
risorsa di calcolo per risolvere problemi

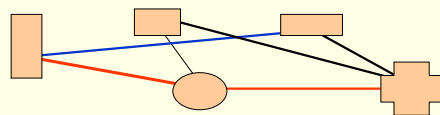
10

un ambiente per il C.D. e' ...



Una **Rete Aziendale**
 composta da
differenti calcolatori
 collegati tra loro tra
reti differenti

Calcolatore Distribuito



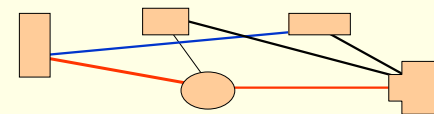
11

un ambiente per il C.D. e' ...



La **Rete** di tutti i
calcolatori
 connessi tra loro
 mediante **Internet**

Calcolatore Distribuito



12

Un ambiente per il C.D. e':

Un sistema di unità processanti **non omogenee**
autonome, indipendenti, geograficamente distribuite
che **sono aggregate**
per risolvere problemi su larga scala
in maniera efficiente

**Risorse condivise
e distribuite**

13

Prima risposta

Calcolatore parallelo

sistema di nodi collegati
da **switch specializzati e
dedicati**
(*tightly coupled systems*)

Sistema ad arch.
distribuita

sistema di nodi collegati
da **reti geografiche**
(*loosely coupled systems*)



La differenza e' nella **rete di connessione**

14

Caratteristiche di una rete di connessione

Tempo per trasmettere un messaggio di n bit

$$t_{comm} = \tau_0 + n \cdot \tau_{comm}$$

Tempo di latenza =
(tempo di start up)=
tempo di comunicazione
di un messaggio nullo

Tempo unitario di trasmissione =
Tempo per trasmettere un
messaggio di 1 bit

15

Velocita' di trasmissione

La **velocita' di trasmissione** di una rete e'
misurata mediante la
Bandwidth (ampiezza di banda) =
velocità di trasferimento di un messaggio
(bit/sec)

Es.: Bandwidth = 30 Mbit/sec



$$\tau_{comm} = \frac{1}{30 \cdot 10^6} \text{sec} = 0.33 \times 10^{-7} \text{sec}$$

16

esempio

	Latenza μ s	Bandwidth Gbit/sec	topologia
Conn. ADSL	1000	< 0.01 (eff. Disponib.)	
LAN Fast Ethernet	100	< 0.1 (eff. Disponib.)	bus
LAN Giga Ethernet	30	1	bus
Infiniband 4x	6	6.5	tree
Myrinet (E)	6	7	switch
QsNetII (R)	3	7	tree
IBM Blue Gene	1 - 5	20	toro 3D

Rete
lenta

Rete
veloce

17

Differenze tra le reti di connessione

	Switch dedicati	Reti LAN (Locali)	Reti WAN (Geografiche)
bandwidth	~ 10 Gb/sec	0.01 - 0.1 Gb/sec	< 0.001 Gb/sec
latenza	< 10 μ sec	10 - 100 μ sec	> 1000 μ sec

18

Seconda domanda:

Che differenza c'è tra gli obiettivi di:

un **Calcolatore Parallelo**



e

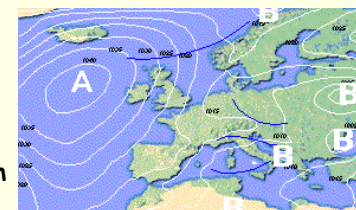


di un **Sistema** per il **Calcolo Distribuito**

19

Esempio: previsioni meteorologiche

Superficie =
20 milioni di Km²
Altezza s.l.m. =
di 20 Km
Discretizzazione =
cubi di lato 100m
(1Km³ = 10³ cubi)



$$20 \times 10^6 \text{ Km}^2 \times 20 \text{ Km} = 4 \times 10^8 \text{ Km}^3$$

$$= 4 \times 10^8 \times 10^3 = 4 \times 10^{11} \text{ cubi}$$

20

Esempio (cont.)

- previsioni per i prossimi 2 giorni (48 ore)
- ogni ora di simulazione in ogni cubo necessarie 10^2 flop

Intera superficie = $4 \times 10^{11} \times 10^2 \times 48 =$
 $\sim 2 \times 10^{15}$ operazioni

PC 1 Gflops

2×10^6 sec
(23 giorni ???)

roadrunner

$\sim 10^2$ sec
(2 minuti !!!)

21

Calcolo parallelo ...al fine di ...

- Ridurre il tempo necessario alla risoluzione computazionale di un problema



Suddivisione del problema



22

Esempio : Progetto SETI

- Progetto per la ricerca di intelligenze extraterrestri
- Dati raccolti dal radiotelescopio di Arecibo
 - 2 anni di raccolta dati (1998-2000)
- Trovare sequenze regolari nelle frequenze all'interno dei segnali raccolti



39 Tbytes =
1100 nastri da 35 Gbyte

23

Le dimensioni del progetto

- ogni nastro da 35 Gbyte e' diviso in oltre 150000 workunit di circa 350 Kb
- ogni workunit contiene i segnali raccolti in circa 100 sec su ognuna delle quali vengono eseguite FFT con diversi campionamenti ed analisi statistiche per circa 280 miliardi flops (circa 40 ore su Pentium II 500 MHz)
- totale 165×10^8 blocchi di circa 100 sec., per un totale di 460×10^{19} flops (quasi 1 mese su Roadrunner !!)

24

Altre caratteristiche

Le workunits sono indipendenti



Predisposizione al parallelismo

Non c'è urgenza ad avere il risultato



Assenza di vincoli temporali

25

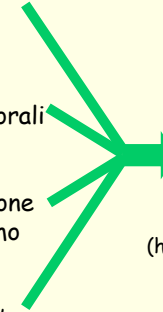
Driving forces

Grandi dimensioni

Assenza di vincoli temporali

Predisposizione al parallelismo

Disponibilità di risorse



IDEA !!

Utilizzare i personal computer nel mondo connessi ad Internet e lasciati inutilizzati
(<http://setiathome.ssl.berkeley.edu/>)

26

Come partecipare?

- collegarsi a (<http://setiathome.ssl.berkeley.edu/>)
- scaricare e installare un apposito software sul PC
- Quando il PC è inutilizzato, in maniera analoga ad uno screensaver, il software si attiva
 - Scarica i dati (~350KBytes) dal server
 - li analizza (circa 10 ore CPU su Pentium4 a 2.0 GHz)
 - restituisce al server i risultati, che vengono combinati con quelli degli altri PC
- ~ 5000000 partecipanti → ~ 100 Tflops di media

27

Calcolo distribuito ... al fine di ...

- riutilizzare "efficacemente" risorse hardware e software distribuite geograficamente sul territorio



Aggregazione di risorse



28

Seconda risposta

Calcolatore parallelo

Principale obiettivo:
Riduzione efficiente dei
tempi di esecuzione

Sistema ad arch.
distribuita

Principale obiettivo:
Aggregazione efficiente di
risorse esistenti



La differenza e' nelle **esigenze**
delle **applicazioni** da risolvere

29

Terza domanda:

Che **differenza** c'e' tra l'uso di:

un **Calcolatore Parallelo**

e

un **Sistema** per il **Calcolo Distribuito**



30

uso delle risorse

Calcolatore parallelo:

- risorse di calcolo omogenee
- risorse sotto controllo diretto

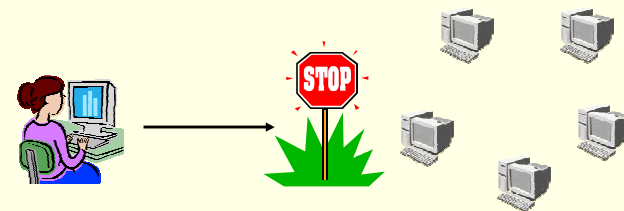


Uso di **librerie di comunicazione** (es. MPI)

Quali sono i problemi nell'utilizzo di un
sistema per il **Calcolo Distribuito**?

31

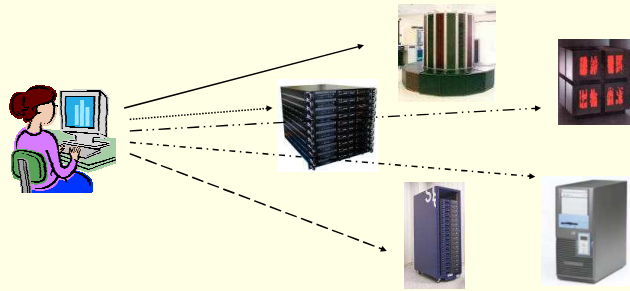
In un ambiente per il C.D. ...



L'utente potrebbe non avere le **autorizzazioni**
per accedere a tutte le macchine

32

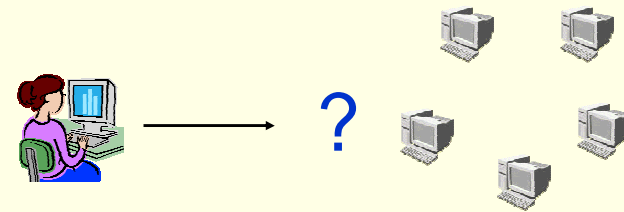
In un ambiente per il C.D. ...



L'utente deve gestire di persona l'**eterogeneita'** delle risorse (sistemi operativi, aritmetica f.p., rappresentazione dati, architettura,...)

33

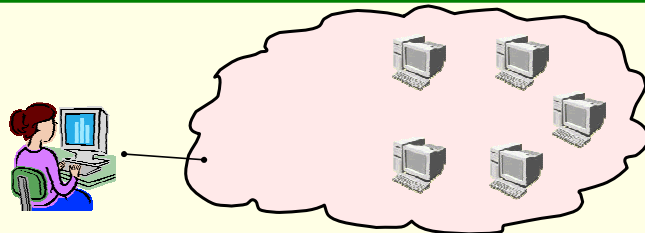
In un ambiente per il C.D. ...



L'utente deve verificare di persona la **disponibilita'** delle risorse e le loro caratteristiche **statiche** (performance) e **dinamiche** (carico di lavoro)

34

Una soluzione



Utilizzare un **ambiente software** tra l'utente e le risorse, in maniera che

- all'utente appaia un **singolo sistema** di calcolo
- risolva i **problemi di eterogeneita', dinamicita' e di sicurezza** dell'ambiente

35

Quindi ...

Uso di un **Sistema per il Calcolo Distribuito**



- risorse non dedicate
- risorse di calcolo non omogenee
- risorse su cui non si ha controllo diretto



Uso di **ambienti software** sofisticati

36

Terza risposta

Calcolatore parallelo

Compiti dell'ambiente sw:
• trasf. dati tra le risorse
(overhead < 5%)

Sistema ad arch. distribuita

Compiti dell'ambiente sw.:
• autenticazione,
• gestione eterogeneita', dinamica, asincronia
• selezione risorse
• trasf. dati
(overhead > 20%)

La differenza e' nel peso dell'ambiente software

37

Quarta domanda:

Che differenza c'e' tra il costo di:

un **Calcolatore Parallelo**



e



di un **Sistema** per il **Calcolo Distribuito**

38

Esempio: IBM BlueGene

L'investimento per la costruzione =
200 M USD ~ 170 M EUR
(senza contare la gestione quotidiana...)

Performance =
450 Tflops

Costo di 1 Tflops = $170/450 =$
0.3 M USD

39

Esempio: SETI@home

L'investimento per i server =
0.5 M USD ~ 0.4 M EUR

performance =
100 Tflops

Costo di 1 Tflops = $0.4/100 =$
0.005 M USD = 4000 EUR !!

40

Quarta risposta

Calcolatore parallelo

↓
Costi hardware
notevoli

Sistema ad arch.
distribuita

↓
Costi hardware
trascurabili

La differenza e' nell'efficienza dell'investimento

41

Parallelo vs distribuito

Calcolatore parallelo

- reti veloci
- risorse limitate
- risorse dedicate e omogenee
- applicazione gestisce le risorse
- costo hardware notevole
- overhead sw sistema < 5%
- presenza di vincoli temporali
- es. IBM Roadrunner
 - 16000 CPU
 - 1 Pflops

Ambiente per il C.D.

- reti lente
- risorse potenzialmente illimitate
- risorse condivise e disomogenee
- ambiente sw.gestisce le risorse
- costo hardware trascurabile
- overhead sw sistema > 20%
- assenza di vincoli temporali
- es. SETI@home
 - 5 milioni CPU
 - 100 Tflops

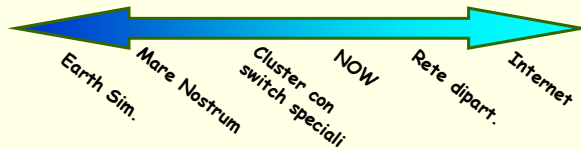
Molte differenze !!!

42

Parallelo vs distribuito

Calcolatore parallelo

Ambiente per il C.D.



Ma non una differenza netta !!!

43

Definizione

Calcolo distribuito

=

Risoluzione efficiente di un problema su un
sistema ad architettura distribuita

↓
Studio delle metodologie di base, degli
algoritmi e conoscenza dei relativi strumenti
software

44

Sinonimi di calcolo distribuito

- network computing
- heterogenous computing
- cluster computing
- meta computing
- ...

Enfasi sui **diversi aspetti** del
Calcolo distribuito

45

FINE LEZIONE

46

Bibliografia e materiale

Generalita' sul calcolo parallelo e distribuito:

- A. Murli - Lezioni di Calcolo Parallelo
- I. Foster - Designing and Building Parallel Programs,
on-line textbook → <http://www-unix.mcs.anl.gov/dbpp/>
- Fox, Williams, Messina - Parallel Computing Works,
on-line textbook → <http://www.netlib.org/utk/lsl/pcwLSI/text/>

- copia dei lucidi
- link internet
- rapporti tecnici e lavori

} distribuiti a lezione
e via e-mail

Per cominciare:

- Dongarra, Meuer, Simon and Strohmaier - High Performance Computing Today - <http://www.netlib.org/utk/people/JackDongarra/PAPERS/hpc-today.pdf>

48

