

PARALLEL HIGH PERFORMANCE COMPUTING
CdS magistrale in informatica

9 – introduzione al calcolo distribuito
Dipartimento di Matematica e Applicazioni
Universita' degli Studi di Napoli Federico II

wpage.unina.it/lapegna

1

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

prima domanda

Che differenza c'è tra:

un **Calcolatore Parallelo**

e

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

2

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

Un calcolatore parallelo e'...

IBM Sierra

Cluster di 4608 nodi di calcolo
Ogni nodo composto da
2 CPU IBM Power9 **24-core**
6 GPU Nvidia Volta

P.P. 125 Pflops

3

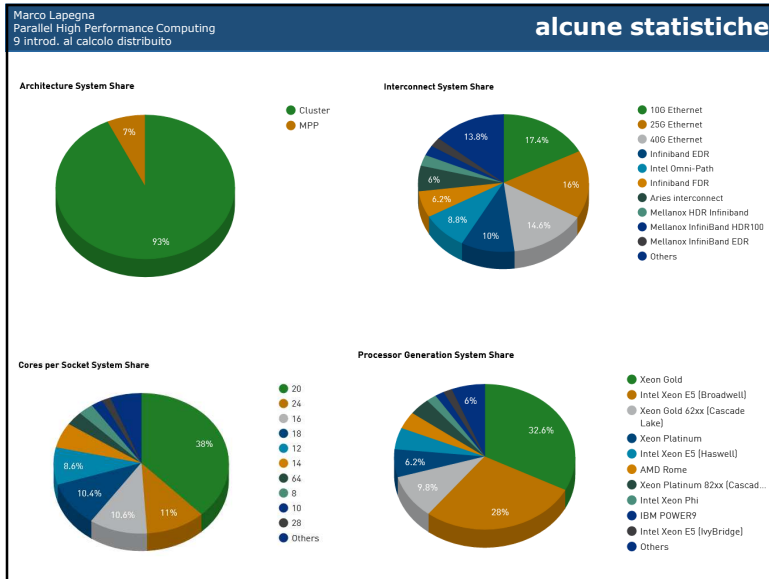
3

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

top 500 list: novembre 2020

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	Supercomputer Fugaku - Supercomputer Fugaku, A64FX 48C 2.20Hz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442,010.0	537,212.0	29,899
2	Summit - IBM Power System AC922, IBM POWER9 22C 3.070Hz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	200,794.9	10,096
3	Sierra - IBM Power System AC922, IBM POWER9 22C 3.16Hz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
4	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.450Hz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
5	Selene - NVIDIA DGX A100, AMD EPYC 7742 64C 2.250Hz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	555,520	63,460.0	79,215.0	2,646

4



5

un calcolatore parallelo e'

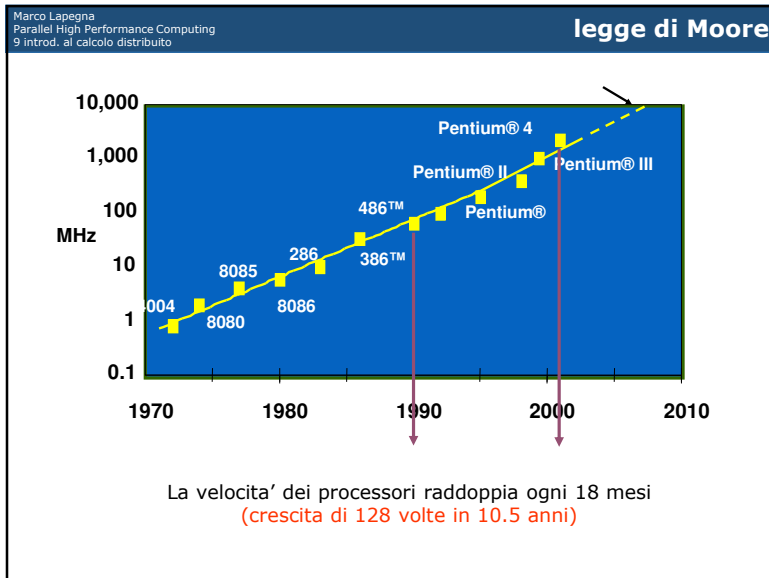
Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

Un sistema di unità processanti **omogenee**
strettamente collegate
che **comunicano**
per risolvere problemi su larga scala
in maniera efficiente

Risorse dedicate e centralizzate

Calcolatore Parallelo

6



7

e il bandwidth delle reti?

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

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	1996	32Mbit/sec
GARR-B	1998	155 - 2500 Mbit/sec
GARR-G	2004	2.5 - 10 Gbit/sec
GARR-X	2014	40 - 100Gbit/sec


Crescita di circa **1000 volte in 10 anni**
(x2 in 9 mesi)

8

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

idea alla base del calcolo distribuito

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

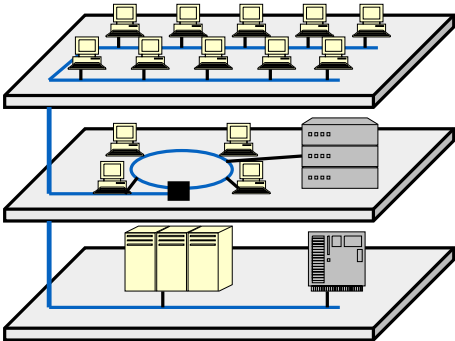


Posso considerare tutti i **calcolatori connessi ad una rete** come una **unica risorsa di calcolo** per risolvere problemi

9

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

un sistema per il calcolo distribuito e':



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

10

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

un sistema per il calcolo distribuito e':



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

11

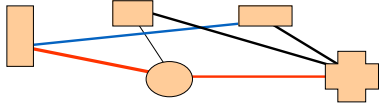
Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

un sistema per il calcolo distribuito 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

Calcolatore Distribuito



12

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

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' nel modo con cui sono collegate le risorse (rete di comunicazione)

13

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

velocita' di comunicazione

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

Es.: Bandwidth = 30 Mbit/sec

↓

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

14

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

esempio

	Latenza ms	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

15

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

differenze

	Switch dedicati	Reti LAN (Locali)	Reti WAN (Geografiche)
bandwidth	~ 50 Gb/sec	0.01 -1 Gb/sec	<0.01 Gb/sec
latenza	< 5 msec	10 - 100 msec	> 1000 msec

↓

• supercomputers

↓

• cluster

↓

• Internet

← Parallelo Distribuito →

16

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito


seconda domanda

Che **differenza** c'è tra **gli obiettivi** di:

un **Calcolatore Parallelo**

e

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



17

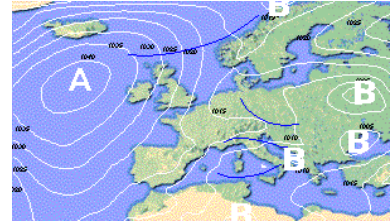
Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

esempio: previsioni meteorologiche

Superficie =
20 milioni di Km²

Altezza s.l.m. =
di 20 Km

Discretizzazione =
celle di lato 100m
(1Km³ = 10³ celle)



↓

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

4×10^{11} celle

18

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

esempio (cont.)

- previsioni per i prossimi 2 giorni (48 ore)
- ogni ora di simulazione in ogni cubo necessarie **10² 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 ???)

supercomputer

↓

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

19

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito


obiettivo del calcolo parallelo

- Ridurre il tempo necessario alla risoluzione computazionale di un problema

↓

Suddivisione del problema

Adatto a problemi dove il fattore tempo è cruciale



20

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

Progetto SETI@home (1999)

Progetto per la ricerca di intelligenze extraterrestri

Ricerca di sequenze regolari di frequenze all'interno dei segnali radio raccolti dal radiotelescopio di Arecibo

Progetto a base volontaria dove ognuno puo' mettere a disposizione il proprio computer quando non viene utilizzato, facendo partire un programma al posto dello screensaver

Migliaia di nastri da 35 GB. ogni nastro da 35 Gbyte e' diviso in oltre 150000 workunit di circa 350 Kb

Completato nel marzo 2020. Ha permesso lo studio della radiazione cosmica di fondo e la sperimentazione di tecnologie software per il calcolo distribuito



21

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

le dimensioni del progetto

ogni workunit contiene i segnali raccolti in circa 100 sec di osservazione in una data frequenza

- i dati sono **ospitati su server** presso l'Univ. della California
- Le **workunit vengono spedite ai client** partecipanti
- i **client eseguono FFT** con diversi campionamenti ed analisi statistiche per circa 3×10^{12} flops
- i **risultati vengono rispediti ai server** a Berkley

- totale 165×10^8 blocchi di circa 100 sec., per un totale di 460×10^{19} flops (quasi 1 mese su un supercomputer !)

Numerosi progetti analoghi

- genome@home
- folding@home (oltre 100 Pflops nel 2016 !!)
- QMC@home
- BOINC

22

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

driving forces

Grandi dimensioni

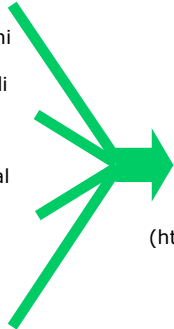
Assenza di vincoli temporali

Predisposizione al parallelismo

Disponibilita' di risorse

IDEA !!

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



23

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito


obiettivi del calcolo distribuito

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

↓

Aggregazione di risorse

Adatto per progetti di grandi dimensioni senza vincoli temporali



24

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

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

25

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito


terza domanda

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

un **Calcolatore Parallelo**

e

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



26

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

uso di risorse di un calcolatore parallelo

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**?

27

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

in un sistema distribuito



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

28

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

in un sistema distribuito

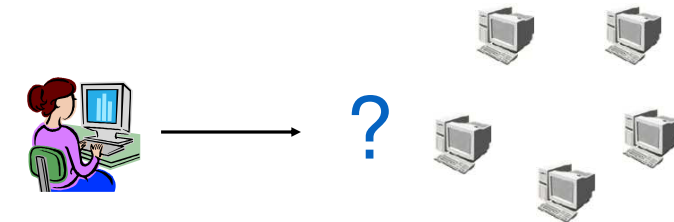


L'utente deve gestire di persona l'eterogeneità delle risorse (sistemi operativi, aritmetica f.p., rappresentazione dati, architettura,...)

29

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

in un sistema distribuito

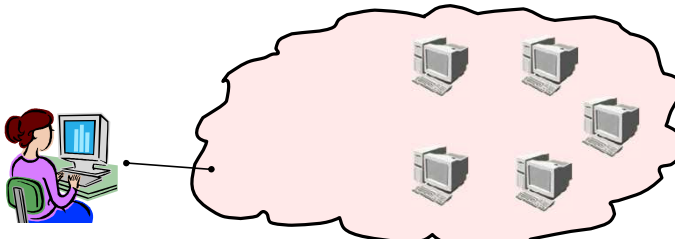


L'utente deve verificare di persona la disponibilità delle risorse e le loro caratteristiche **statiche** (performance) e **dinamiche** (carico di lavoro)

30

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

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, dinamicità e di sicurezza** dell'ambiente eterogeneità

31


Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

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

32

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

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 ruolo dell'ambiente software

33

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito


quarta domanda

Che **differenza** c'e' tra il **costo** di:

un **Calcolatore Parallelo**

e

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



34

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

un confronto

Esempio: IBM BLUE Gene

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 EUR

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 !!

35

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

quarta risposta

Calcolatore parallelo

↓

Costi hardware
notevoli

Sistema ad arch. distribuita

↓

Costi hardware
trascurabili

↓

La differenza e' nell'efficienza dell'investimento

36

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

parallelo vs distribuito

<p>Calcolatore parallelo</p> <ul style="list-style-type: none"> • 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 <ul style="list-style-type: none"> • 16000 CPU • 1 Pflops 	<p>Ambiente per il C.D.</p> <ul style="list-style-type: none"> • 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 <ul style="list-style-type: none"> • 5 milioni CPU • 100 Tflops
--	---

Molte differenze !!!

37

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

parallelo vs distribuito

<p>Calcolatore parallelo</p>	<p>Ambiente per il C.D.</p>
-------------------------------------	------------------------------------

Ma non una differenza netta !!!

38

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

definizione

Calcolo distribuito
 =
 Insieme delle metodologie di base e
 degli strumenti software per la
Risoluzione efficiente di un problema su un
sistema ad architettura distribuita

39

Marco Lapegna
Parallel High Performance Computing
9 introd. al calcolo distribuito

alcuni sinonimi e/o variazioni

- network computing
- cluster computing
- grid computing
- cloud computing
- edge computing

Enfasi sui **diversi aspetti** del **Calcolo distribuito**

40