

# Sviluppo di una infrastruttura HPC (High Performance Computing) e GRID per il calcolo scientifico in un IRCCS

Bava Michele<sup>1</sup>, Dadamo Pio<sup>2</sup>, Cacciari Domenico<sup>3</sup>, Cozzini Stefano<sup>4</sup>

<sup>1</sup>Clinical Engineer, IT Ast. Manager - Servizio Informativo IRCCS "Burlo Garofolo" Trieste

<sup>2</sup>2° y. PhD Student - DEEI – Dipartimento di Elettronica Elettrotecnica ed Informatica - Università di Trieste

<sup>3</sup>S.C.U. Genetica Medica – IRCCS "Burlo Garofolo" Trieste

<sup>3</sup>Servizio Informativo IRCCS "Burlo Garofolo" Trieste

<sup>4</sup>Direttore E-Lab – Laboratory for e-Science – SISSA - Trieste

# Lo scenario

- L'IRCCS “Burlo Garofolo”
- La ricerca clinica
- Il calcolo scientifico
- Necessità e bisogni
- Le soluzioni adottate
- Il progetto
- Gli sviluppi
- Conclusione

# Lo scenario

- Sempre maggiore necessità di risorse di calcolo a disposizione del mondo scientifico e della ricerca
- Ricerca clinica e assistenza medica (esigenze di calcolo, di storage, di reperimento informazioni, ecc...):
  - Soluzioni per la diagnosi e la cura delle malattie (malattie multifattoriali)
  - Studi epidemiologici (analisi e studio di grosse moli di dati)
  - PACS regionali (immagini, referti)
- La Sanità inizia ad investire in soluzioni tecnologiche e in risorse umane partendo dalla ricerca per giungere ad una ricaduta pratica nella diagnosi e cura delle malattie
- Manca però ancora in Sanità un know-how adeguato su questi sistemi, tecnologie e servizi -> collaborazioni con altri enti di ricerca (Università, laboratori, centri ricerca).

# L'IRCCS “Burlo Garofolo”

- E' un ospedale di alta specializzazione e di rilievo nazionale nel settore pediatrico ed in quello della tutela della maternità e della salute della donna.
- Collabora strettamente con l'Università di Trieste ospitandone le Cattedre di Pediatria, Ostetricia e Ginecologia, Igiene e Genetica e le relative Scuole di Specializzazione.
- Impiega circa 800 persone, tra le quali una cinquantina di giovani ricercatori, borsisti e contrattisti.
- In quanto IRCCS, persegue finalità di ricerca nel campo biomedico ed in quello dell'organizzazione dei servizi sanitari, di innovazione nei modelli d'assistenza, di formazione e di trasferimento delle conoscenze.

# L'IRCCS “Burlo Garofolo”

- Aree di **eccellenza assistenziale** sono costituite:
  - dalle cure multispecialistiche alle malattie rare e complesse del bambino;
  - dagli approcci mini-invasivi in chirurgia pediatrica e ginecologica;
  - dall'assistenza a gravidanze ad alto ed altissimo rischio per patologie materne, con una assistenza neonatale che consegue risultati di assoluto rilievo (mortalità neonatale inferiore al 2 per mille) in un contesto di massima attenzione ai bisogni del neonato e dei familiari.

# L'IRCCS “Burlo Garofolo”

- L'attività di ricerca, coordinata dalla Direzione Scientifica, trae obiettivi e stimoli dall'attività clinica così come da quella epidemiologica, e mira a trasferire il prodotto di nuove conoscenze, nuove tecnologie e modelli assistenziali all'insieme del servizio sanitario regionale, nazionale ed alla comunità scientifica internazionale.
- L'attività di ricerca si effettua nelle aree della salute riproduttiva, della medicina materno-fetale, della neonatologia, delle malattie geneticamente trasmesse e delle malattie croniche e tumorali con esordio in età pediatrica, della chirurgia pediatrica generale e specialistica, delle neuroscienze dell'età evolutiva, della ginecologia, e si estende alla epidemiologia dei fattori ambientali, infettivi e degli stili di vita rilevanti per la salute di madri, bambini e adolescenti ed alla ricerca operativa sui servizi sanitari.

# L'IRCCS “Burlo Garofolo”

- L'attività scientifica si traduce in un centinaio di pubblicazioni all'anno su riviste di altissimo prestigio scientifico nei rispettivi campi, quali, *Lancet*, *New England Journal of Medicine*, *British Medical Journal*, *Proceedings of National Academy of Science*, *Pediatrics*, *American Journal of Obstetrics and Gynecology*, *Human Reproduction*, ecc.).
- Il risultato della attività scientifica è particolarmente buono se rapportato alle ridotte dimensioni dell'Istituto ed al numero dei ricercatori attivi (120, di cui 2/3 impegnati prevalentemente in attività assistenziali).
- Il prodotto di ricerca si caratterizza per le sue ricadute sul Servizio Sanitario Nazionale: circa l'80% delle pubblicazioni riferisce conoscenze e tecnologie applicabili a breve/medio termine alla diagnostica, alla cura ed alla prevenzione.

# L'IRCCS “Burlo Garofolo”

- L'Istituto partecipa a **progetti finanziati dalla UE** quali Blueprint for Action (allattamento al seno), HOPE (determinanti precoci dell'obesità), EMBIC (meccanismi di controllo dell'impianto embrionale), e PHIME (esposizione prenatale a metalli ed in particolare al mercurio).
- Il Burlo Garofolo vanta una cospicua rete di **collaborazioni internazionali**, con le Agenzie delle Nazioni Unite in primo luogo e con una vasta rete di istituzioni e società scientifiche sia nei paesi industrializzati che in via di sviluppo.
- Dal 1992, il Burlo Garofolo è riconosciuto quale *Collaborating Centre* dell'OMS per la salute materno-infantile e opera, attraverso l'Unità per la salute internazionale del bambino, sul versante della cooperazione internazionale per lo sviluppo e la valutazione di tecnologie appropriate per l'assistenza alla gravidanza, al parto, al periodo neonatale, all'infanzia e all'adolescenza;

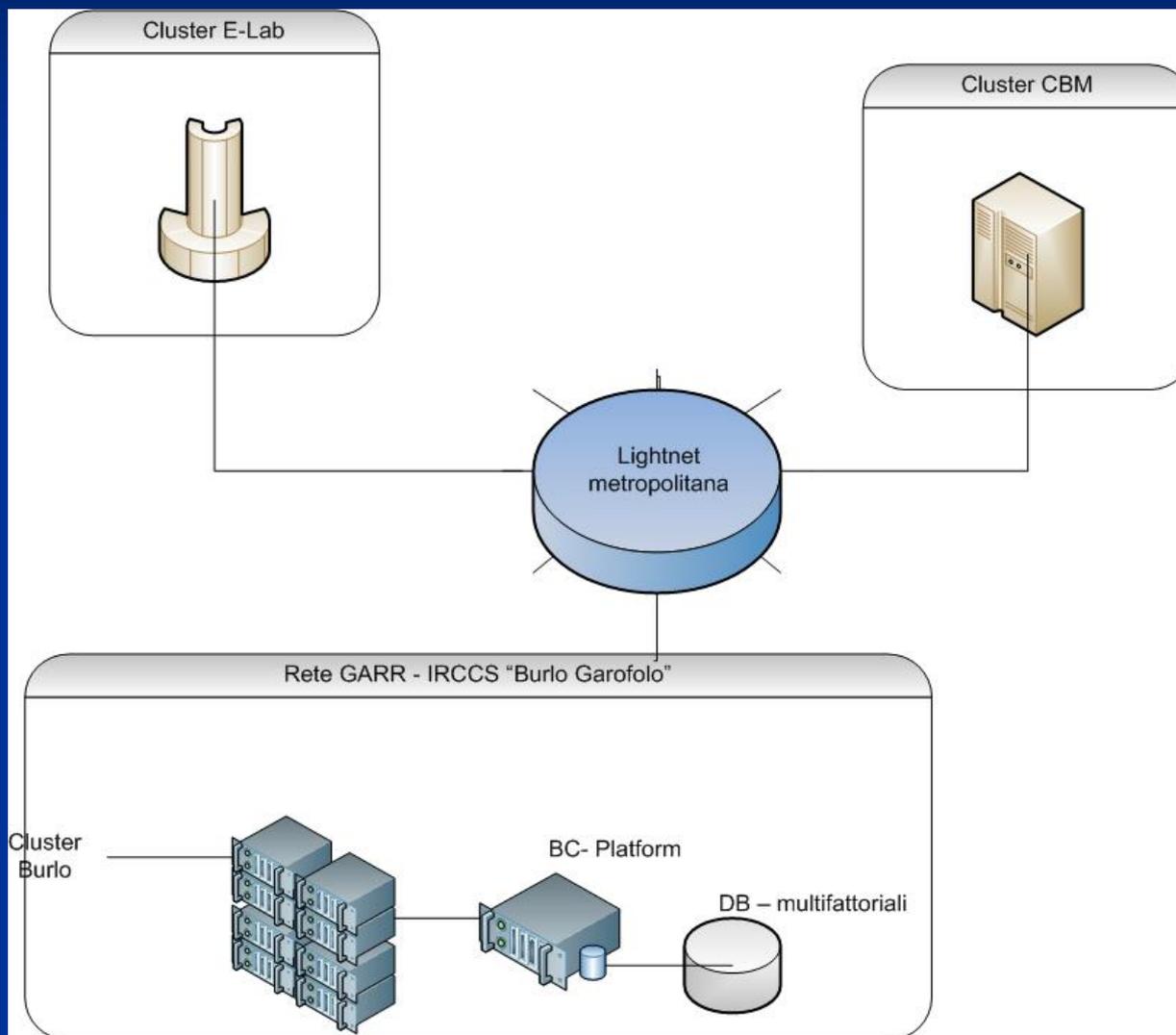
# Il Sistema informativo dell'IRCCS

- Parte della S.C. Ingegneria Clinica e Acquisizione Tecnologie
- Ci occupiamo di:
  - interventi, manutenzioni, servizi di supporto, gestione applicativi, gestione e sviluppi delle reti
  - ricerca (telemedicina, portale integrato di servizi, IT security, sistemi di gestione della conoscenza aziendale, sistemi e strumentazione innovativa per il trattamento di disturbi pediatrici, NIRS e modellazione 3D)
  - sviluppi (collaborazioni con le SC dell'Ospedale, con UniTS, con associazioni ed Enti professionali e di ricerca – AIIC, @ITIM, GNB, SISSA, CBM, ISF, AGMEN; partecipazione a progetti RC, ministeriali o universitari – Start Cup 2009)
- **Attenzione alle esigenze delle strutture interne dell'Istituto e dei ricercatori**
- Apertura a collaborazioni con altre Aziende Sanitarie e con altre realtà (scientifiche e “business”) dell'area scientifica-tecnologica triestina

# Trieste, città della scienza

- Estrema concentrazione di realtà scientifiche e di ricerca sul territorio (il Burlo è una di queste!)
- Esistono convenzioni e MOU in atto con l'Università di Trieste (in particolare DEEI e Dipartimento dei Materiali), con il CBM (Centro di Biomedicina Molecolare), con la SISSA (E-Lab in particolare).
- Collaborazioni con l'Area di Ricerca (AREA Science Park):
  - Insiel S.p.A.
  - ItalTBS S.p.A.
  - SPIN s.r.l.

# Infrastruttura



# La ricerca clinica

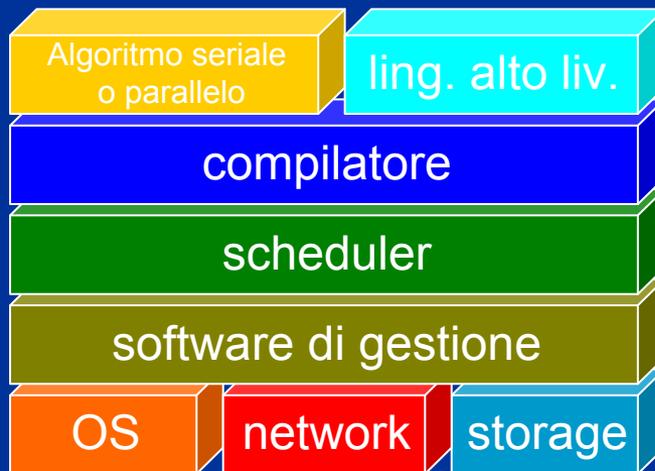
- Sono elencati alcuni dei più recenti ambiti di ricerca con risultati di rilievo da un punto di vista assistenziale:
  - Uso dell'antibiotico nei bambini con otite acuta
  - Nuovi strumenti di aggiornamento per i pediatri generalisti
  - La gravidanza ad alto rischio
  - Nuovi protocolli per il trattamento della fibrosi cistica
  - Risonanza magnetica fetale e chirurgia pediatrica
  - Depressione, post partum e violenza
  - **Il parco genetico del Friuli Venezia Giulia (in particolare lo studio delle malattie multifattoriali o complesse)**

# Il calcolo scientifico

- E-science: nuovi aspetti della scienza che vengono trattati con risorse di calcolo intensive
  - High performance computing
  - Parallel computing/computers
  - Clusters
  - Grids

# Il calcolo scientifico - cluster

- Si tratta di macchine individuali combinate in un unico sistema per mezzo di un software e di una rete dedicata
- E' formato da un master node e da più cluster nodes su cui reiterare le installazioni a seconda dell'uso che si farà



Infrastruttura software di un cluster

# Il calcolo scientifico: grid

- Ambienti di calcolo all'interno dei quali le applicazioni hanno bisogno di accedere in maniera veloce ed affidabile a grandi quantità di dati geograficamente distribuiti.
- Conoscenza e risorse sono delocalizzate
  - I laboratori o centri di ricerca sono distribuiti
  - I dati scientifici sono distribuiti
- Si tratta quindi di molte risorse (computers) geograficamente distribuiti, collegati da una rete e supportati da un software (middleware) che tiene unite tutte le risorse
- Trasparenti per l'utente

# Il calcolo scientifico: grid

- Le applicazioni del grid computing:
  - Computazionali: simulazioni intensive (meteo), modellazione ingegneristica
  - “Data intensive”: analisi di dati sperimentali; analisi di dati remoti provenienti da sensori; analisi e storage di immagini
  - Collaborazioni distribuite: strumentazione online condivisa, radiodiagnostica, ecc...

# Necessità e bisogni

- Le malattie multifattoriali o complesse (diabete, ipertensione, obesità, infarto miocardio, etc.) dipendono sia da fattori genetici che da fattori ambientali (fumo, abitudini alimentari, stili di vita, etc.). Approfondire le conoscenze sui fattori genetici risulta fondamentale per chiarire l'eziopatogenesi di queste malattie ed aprire nuovi orizzonti terapeutici. Lo studio di popolazioni isolate ha costituito uno dei mezzi per l'individuazione dei geni responsabili di malattie monogeniche, ereditate come caratteri mendeliani semplici, quali ad esempio la talassemia.

# Necessità e bisogni

- In una popolazione isolata, secoli di endogamia hanno limitato il pool di geni trasmessi da una generazione all'altra, hanno favorito l'instaurarsi di alti livelli di omozigosità e l'insorgere di malattie rare recessive. Analogamente e tramite lo stesso meccanismo, si può ipotizzare che anche il numero di geni coinvolti in una patologia poligenica sia più basso in un isolato genetico rispetto ad una popolazione più ampia e geneticamente più eterogenea per la presenza di componenti di origine diversa. Molte malattie comuni, quali ad esempio l'ipertensione e il diabete, sono patologie poligeniche in cui all'effetto dell'ambiente si somma una predisposizione genetica più o meno significativa. Lo studio dei fattori genetici e ambientali che determinano la predisposizione a tali patologie è quindi facilitato dallo studio degli isolati genetici, infatti anche la componente ambientale, quali ad esempio il clima e la dieta, è molto più omogenea e tale è rimasta per secoli. Poiché però il numero di geni è grande e quelli identificabili in ciascun isolato saranno pochi, sarà necessario lo studio di diversi isolati genetici per identificare i molti geni responsabili di patologie comuni in popolazione più eterogenee.

# Necessità e bisogni

- L'analisi genetica di queste popolazioni isolate è la componente principale del “Parco Genetico del Friuli-Venezia- Giulia” un progetto che ha come obiettivo la realizzazione di una Biobanca Regionale corredata di dati clinici, storici, ambientali, e biologici, di sei paesi del Friuli Venezia-Giulia che per motivi geografici, storici e/o linguistico-culturali sono rimasti isolati per molto tempo.
- Scopo primario del progetto è fotografare e studiare una serie di malattie "multifattoriali", come l'infarto, l'ipertensione, il diabete, l'osteoporosi, l'obesità, la menopausa precoce, le patologie neurologiche infantili, la dislessia.

# Necessità e bisogni

- Ogni volontario si è sottoposto senza alcuna spesa a uno screening completo: esami-base del sangue e delle urine, una visita di medicina generale con misurazione di peso, altezza e pressione arteriosa, una visita cardiologica con elettrocardiogramma, misurazione della densità ossea, una visita odontoiatrica, un test del gusto con compilazione di un questionario da parte del medico, una visita audiologica e una visita neurologica con questionario. Si stanno archiviando i dati di circa 4mila persone.

# Necessità e bisogni in particolare

- Da ciascun isolato genetico si riesce a derivare un pedigree che riporta i dati genetici relativi ad un certo paese/comunità
- L'analisi di questo pedigree permette ai ricercatori di collegare una zona cromosomica con più fenotipi, in modo da capire meglio l'eziologia di alcune malattie.
- Tutti questi dati devono essere incrociati considerando quindi il pedigree, le informazioni di carattere genetico (di ciascun soggetto) più le informazioni sulle malattie (informazioni fenotipiche).

# La soluzione attuale

- Attualmente il Burlo ha a disposizione un piccolo cluster interno composto da una parte proprietaria (BC-Platform) più la possibilità di eseguire i job necessari all'attività scientifica presso i nodi del CBM o della SISSA.
- La gran parte dei calcoli viene eseguita presso i due “provider”
- A BC-Platform è stato collegato un nodo che permette di eseguire una parte di calcoli. La soluzione è comunque modulare e permette di aggiungere altri nodi.

# La soluzione attuale

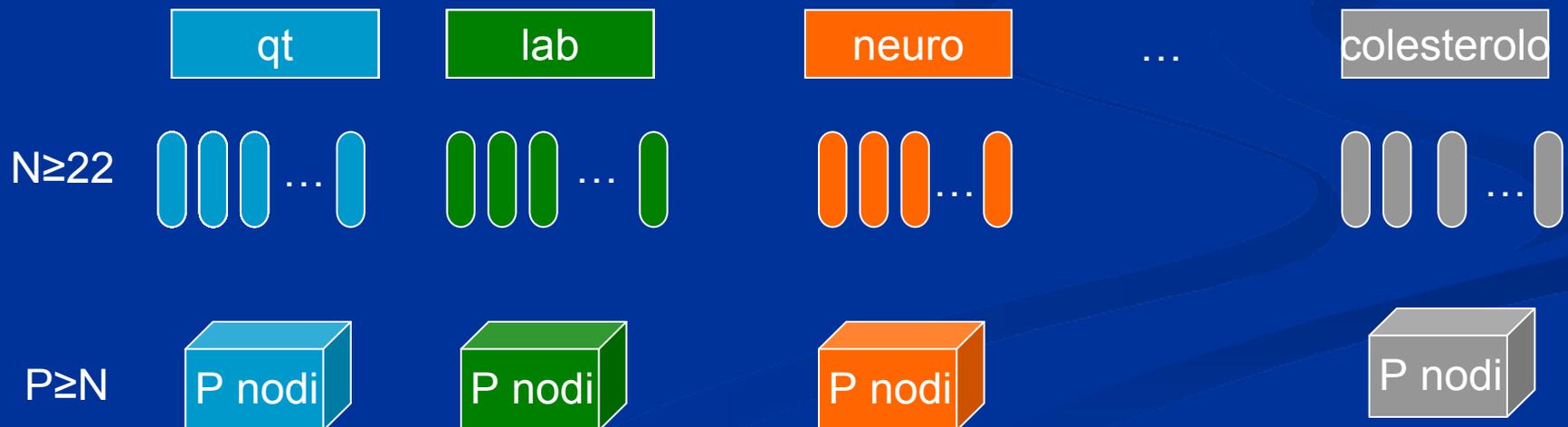
- In BC-Platform, per rispondere alle esigenze del progetto, sono state implementate delle tabelle rispettivamente relative a:
  - Pedigree: tabella con codice famiglia, soggetto, madre, padre e sesso che si estende per quanto è grosso il paese (o isolato genetico) in forma di albero genealogico
  - Informazione genetica: tabella con codice soggetto e codice marcatore (x2 allele 1 & 2) in cui per ciascun soggetto vi sono le informazioni relative ad 1 milione di marcatori “sparsi” sul genoma da cui sono raccolte le informazioni relative alle basi del DNA corrispondente al marcatore stesso
  - Tante tabelle “fenotipiche” relative alle varie informazioni di carattere clinico che sono state estratte e che si vogliono incrociare (es. tabella cardiaca con codice soggetto e dati cardiologici tipo ecg; tabella laboratorio analisi, tabella neurologica, ecc...)

# La soluzione attuale

- Con BC-Platform, per ciascun fenotipo, eseguo una query specifica che ad esempio mette in relazione una zona cromosomica con il fenotipo (p. es. allungamento del tratto QT dell'ecg).
- Tale analisi, per ciascun fenotipo va fatta sui 22 cromosomi e per velocizzare l'analisi posso analizzare i singoli cromosomi in parallelo;
- I primi cromosomi, i più grandi, possono essere a loro volta spezzettati in pezzi più piccoli e dati "in pasto" a diversi nodi/cpu;
- Allo stesso modo il pedigree può essere spezzato in pezzi più piccoli e così diventa più articolata (e pesante) la query per collegare tutti i dati;
- Infine occorre fare un controllo preventivo sui dati detto di coerenza mendeliana che impegna ancora risorse di calcolo
- Più grande il pedigree maggiore il contenuto informativo

# La soluzione attuale

- L'analisi per ciascun fenotipo sui  $N \geq 22$  cromosomi può essere condotta su più fenotipi in modo parallelo
- In linea teorica potremmo gestire in parallelo (su un cluster o su una griglia) ciascun fenotipo parallelizzando su più nodi i fenotipi
- Risorse di calcolo con molta RAM per ciascun nodo in modo da avere pedigree più grandi (ovvero più informativi)



# La soluzione attuale

- In BC platform non c'è una comoda gestione delle priorità
- BC platform è multi-utente permettendo a ciascun gruppo di studio/ricerca di inserire i dati e di analizzarli in modo comodo e relativamente semplice
- L'output che si ottiene, di solito in forma visuale, riporta la probabilità che su quel marcatore c'è la variante genetica responsabile di quel fenotipo
- I tipi di analisi che si possono fare sono molto diversi, ovvero ci sono diversi software che girano sul cluster
- Su bc-platform preparo i vari job secondo i vari software che intendo utilizzare (che scelgo a seconda del pedigree)
- Su ogni nodo del cluster installo i software (simwalk 2, merlin, solar, ...)
- A seconda delle code che imposto ho job che utilizzano un numero di nodi variabile e un diverso quantitativo di RAM per ciascun nodo

# La soluzione attuale

- I bioinformatici del Burlo preparano anche DB e code personalizzate da inviare ad altri cluster (CBM, Sissa)
- I dati del DB sono sempre relativi a genoma, fenotipo e pedigree
- Attraverso degli script fatti in Perl/DBI si preparano i dati in modo da essere inviati al cluster
- Sul cluster è possibile impostare dei sistemi di priorità personalizzabili
- Infine, dopo il calcolo, sono stati fatti anche degli script personalizzati per il parsing dell'output
- Tutto questo viene fatto attraverso una interfaccia da riga di comando direttamente connessi con il cluster

# Prospettive

- In BC platform va inserita una gestione delle priorità che attualmente non c'è
- Non è customizzato sulle esigenze dell'utente
- Interfaccia facile da usare (soprattutto per una gestione con più team multidisciplinari)
- Poca possibilità di piazzare i propri script fatti in Perl o in R
- Profiling (misura dei tempi impiegati ai vari livelli del codice), tecniche di ottimizzazione (utilizzo della cpu, della memoria, dei programmi) del codice e benchmarking (cpu, memoria, collegamento nodi)

## ■ BC Platform

- proprietario
- multiutente
- poco configurabile
- gestione del DB e dell'interfaccia utente agevoli
- costoso



## ■ Soluzioni "home made"

- customizzate
- script personalizzati
- gestione delle code e delle priorità
- maggiore controllo delle risorse
- open



## ■ Know-how SISSA/E-Lab

- sviluppo infrastruttura, aggiungendo a BC-Platform nodi "open"
- alcuni sw non sono ottimizzati, sono scritti male e possono essere ottimizzati nell'uso delle risorse di calcolo a disposizione
- sviluppo della GRID

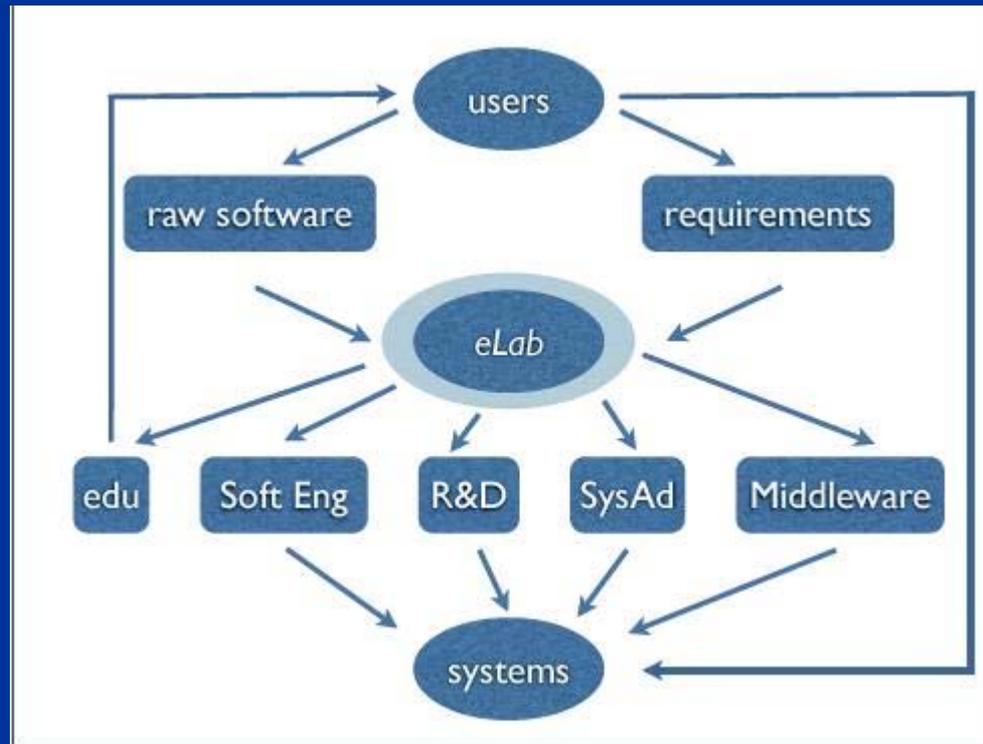
## ■ Sviluppi

- collaborazione tra enti ricerca diversi
- miglioramento diagnosi malattie complesse
- know-how interno al Burlo
- potenza di calcolo da rivendere
- primo esempio in Sanità regionale (Burlo capofila di eventuali progetti)



# Strumenti

- E-Lab supporta lo sviluppo del Cluster e quello della GRID fornendo consulenza e formazione a risorse del Burlo (MOU)



# I passi del progetto

- Riunioni con la SCU Genetica Medica per:
  - Capire i requisiti computazionali (elevati)
  - Capire il grado di conoscenze IT del gruppo (elevato – bioinformatici)
  - Capire fino a che punto conoscono la propria applicazione e/o il loro studio (elevato – occorre ottimizzare quel che fanno già)
  - Capire i requisiti dell'applicazione o delle applicazioni in modo da dimensionare il progetto
    - Potenziare il cluster
    - Sviluppare la Grid
  - Proporre una soluzione ottimizzando gli algoritmi e l'integrazione tra la parte proprietaria e la parte “open” che si vuole implementare
  - Benchmarking

# I passi del progetto

- Step 0: eventuale “check” sulla consapevolezza (1) dell’utilizzo di risorse HPC o GRID da parte degli utenti e analisi delle (2) richieste computazionali
  - (1) Consapevolezza elevata
  - (2) Necessità di molti nodi e di elevati quantitativi di RAM
- Step 1: sviluppo tecnico dell’infrastruttura (scelta tra le varie tipologie di cluster o di grid)

In progress

- Step 2: procedure di benchmarking e valutazione dell’efficienza della soluzione proposta
- Step 3: messa in produzione e test finale
- Step 4: avvio



# Il progetto: lo sviluppo

- Vi sono tutta una serie di problematiche da affrontare per trovare una soluzione:
  - Legate al tipo di applicazione/studio (algoritmo seriale, parallelo, parallelizzabile, ecc...)
  - Legate ad eventuali modifiche da apportare allo studio (ad esempio aspetti legati al modello oppure al campionamento dei dati)
  - Fattori umani (buona volontà e collaborazione; disponibilità di risorse)
  - Disponibilità di strumenti e risorse tecnologiche opportune
  - Quale e quanto parallelismo e a che livello (a livello della cpu, dei nodi, del cluster, della grid)?
  - Quale hardware, quanti nodi (multiprocessore, multicore)?

# Il “mini” cluster

- Sviluppo del cluster interno partendo dalla soluzione proprietaria
- Aggiunta di nodi
- Integrazione (se possibile) della parte proprietaria (gestisce bene il db) con la parte “open” (svincola dal vendor e permette la creazione di script su misura e ottimizzati)
- Ottimizzazione del codice (a seconda di che cluster viene messo su)
- Tuning e benchmarking
- Start! (meglio init!)
- Apertura ad altre SC e/o ad altri utenti

# Il “mini” cluster



# Strumenti

SYSTEM FEATURES - ClusterDocumentation - Windows Internet Explorer

http://www.democritos.it/cluster-wiki/index.php/SYSTEM\_FEATURES

Google

Google

Effettua la ricerca

Segnalibri

Controllo

Traduci

Compilazione automatica

Entra

Preferiti

Siti suggeriti

Raccolta Web Slice

GARR - Italian Academic ...

SYSTEM FEATURES - Cl...

Pagina

Sicurezza

Strumenti

13

Compiling codes

Using Lenses

Using PBS/Torque

Cancel queues

Resolving errors

Site Admin info

type of tools

SSH Tunneling

application and libraries

Quantum Espresso

Name

DMC

PGD/GIS

INP

CPD

Climate Packages

QTC

Grid@Trieste

MIU@EUI

glite Local installations

Guides

GridSeed@EUI

training

Training Activities

Training@GridSeed

edit

help on cluster

Recent changes

Random page

search

Go

Search

tools

what links here

Related changes

Upload file

Special pages

Printable version

Permanent link

**HARDWARE** [edit]

Cluster	Briareo	Cerbero	Borg
Model	IBM 1300 Cluster	Opteron Celestica/Sun/Supermicro Onyxia	Opteron Machine
Server Node	IBM x342 (2 PIII 1.4 GHz) @	Opteron @	Opteron @
Computing nodes	28 IBM x330	88	7
Number of PEs	28x2 PIII 1.4 GHz	14x2 AMD-246 2.0 GHz 24x2 AMD-246 2.6 GHz 12x4 AMD-275 2.2 GHz 8x2 Xeon 3.0 GHz (64 bit)	7x4 AMD Opteron processor- 2405.519 MHz (blades)
L2 Cache	512 Kb per CPU full speed	1024 Kb per CPU full speed 1024 Kb per CPU full speed (64)	1024 Kb per CPU full speed
RAM	28 GB (1 GB/node)	214 GB (4 GB/node)	54.8 GB (7.8 GB/node)
Disk space	1.3 TB	11 TB	177.2 GB
Internal Network	Myricom LAN Card	Gigabit and Myricom LAN Card	Gigabit and Infiniband LAN Card

**SOFTWARE** [edit]

Cluster	Briareo	Cerbero	Borg
Operating System	Linux (RH 7.3)	Linux (White box, version 4)	Linux (White box, version 4)
Kernel Version	2.4.20-30amp	2.6.14	2.6.21 3-diskless for blade nodes
Cluster Software	CSM Version 1.3.0.20	pxe-ks+C3	pxe-ks+C3
Parallel Filesystem	GPFSS	NFS	NFS
Myrinet	GM Version 1.0.3	GM Version 2.0.23	-
MPI Library	MPICH-GM Version 1.2.5.9	MPICH	OPENMPI
Available compilers	Portland Group Fortran F90, C, C++, GNU compilers; Intel F90 compiler	Portland Group Fortran F90, C, C++, GNU compilers; Intel F90 compiler and g95 compiler	Portland Group Fortran F90, C, C++, GNU compilers; Intel F90 compiler and g95 compiler
Queue System	PBS+MAUI (parallel jobs)	TORQUE+MAUI (parallel jobs)	No queue system, its a test/development cluster
Parallel libraries	MPI	MPI	MPI

**FILESYSTEM AND DISK SPACE** [edit]

	Briareo	Cerbero	Borg
home filesystem	33 Gb (as /home mounted via nfs)	2.7 Tb (as /u mounted via nfs)	9.2 Gb (as /u mounted via nfs)
global scratch filesystem (viewable from all nodes)	130 Gb (as /scratch mounted via gpx)	933Gb (as /scratch mounted via nfs)	100 Gb (as /scratch mounted via nfs)
local scratch filesystem (on each compute node)	9.7 Gb	60 Gb (as /local_scratch)	blade nodes are diskless

# La grid

- Per passare alla griglia occorreranno importanti cambiamenti sull'attuale modo di eseguire i calcoli e fare l'analisi
- Approcci per il porting delle applicazioni della genetica su grid
  - SMP-multicore/MPI-multiprocessor
  - Client/server (egee)

# La grid

- Client/server (egee)
  - Ci troviamo in presenza di query sui dati poco accoppiate (possiamo far eseguire fenotipi diversi in parallelo)
  - Si possono inviare N job (a N nodi) indipendenti e coordinarli attraverso una specifica architettura client/server
  - Ogni nodo esegue un fenotipo (p. esempio)
  - Le comunicazioni/gli scambi tra i clients sono coordinate dal meccanismo client/server
  - All'utente rimane l'onere di usare (attraverso script) in modo efficiente e parallelo ciascun nodo

# La grid



# La rete GARR e la MAN

- La rete del GARR, LightNet e l'Università sono partner preferenziali per lo sviluppo dell'infrastruttura HPC e soprattutto GRID.
- Entrare in progetti tipo [GRID@TRIESTE](#) (ICTP, SISSA, Sincrotrone Trieste, INAF e Università di Trieste attraverso il DMI) nato nel 2005 con l'obiettivo di promuovere lo scambio di esperienze e know-how acquisiti all'interno di ciascun Istituto sulla tecnologia e le problematiche GRID e di realizzare una infrastruttura GRID metropolitana a beneficio di tutta l'utenza scientifica afferente agli Istituti che partecipano al progetto.

# Conclusioni

- Sanità: scenario sempre più complesso (non è una novità...)
- Necessità di supportare sempre più studi e ricerche nell'ambito e con gli strumenti dell'e-science
- Maggiori investimenti in risorse (tecnologiche ed umane)
- Più ricerca e formazione sulle nuove tecnologie
- Partnership tra enti di ricerca impegnati in settori diversi: know-how incrociato
  - Possibilità di nuovi ambiti di ricerca e sviluppo
- Partire dalle esigenze degli utenti
- Rete del GARR riferimento ideale per uno sviluppo scientifico e tecnologico della Sanità

# Riferimenti

- [www.burlo.trieste.it](http://www.burlo.trieste.it) (per le notizie relative al progetto degli isolati genetici)
- [www.democritos.it](http://www.democritos.it) (il nostro partner E-Lab)
- Progetto IRCCS@GARR
- Materiale della “Advanced School In High Performance and Grid Computing” – ICTP Trieste
- <http://www.grid.ie/mpi/wiki>
- EGEE: enabling grids for e-science
- <http://glite.web.cern.ch/glite/documentation/default.asp>