




Molte reti per una Grid

La rete del futuro che condivide
risorse di calcolo e memoria.

di Mirco Mazzucato e Vincenzo Napolano



Centoquaranta centri di calcolo di trentatré paesi diversi, milioni di gigabyte immagazzinati ed elaborati ogni anno, migliaia di scienziati coinvolti nella gestione e nell'analisi dei dati, che già oggi possono disporre facilmente della potenza di calcolo di decine di migliaia di calcolatori coordinata e moltiplicata a fini scientifici. È la Grid di Lhc, sviluppata per consentire a migliaia di ricercatori di diversi paesi di analizzare i prodotti di milioni di collisioni prodotte dall'acceleratore, una volta che sarà entrato in funzione. Grazie a questo lavoro i fisici sperano di scovare l'origine della materia oscura e il supermondo, le tracce dello sfuggente bosone di Higgs e tentare di comprendere meglio le leggi della fisica all'origine dell'Universo e il Big Bang.

Per questo sarà necessaria l'analisi dei 15 milioni di gigabyte di dati, prodotti ogni anno da Lhc, pari a una torre di cd alta 20 chilometri, ovvero come 60 torri Eiffel una sull'altra.

La gestione e l'analisi di questa immensa mole di dati avverrà attraverso le migliaia di nodi di calcolo della Lhc-Grid sparsi su tutto il pianeta e organizzati in diversi livelli. Gli undici del primo "livello" (Tier 1), tra cui il Cnaf (il centro per la ricerca e lo sviluppo delle tecnologie informatiche dell'Infn) di Bologna, riceveranno direttamente dal Cern i dati prodotti dall'acceleratore, per elaborarli, analizzarli, archivarli e smistarli successivamente agli altri centri sul pianeta.

La Grid di Lhc a oggi è l'unica ad avere un'estensione mondiale, con siti nei diversi continenti, anche se non è l'unica griglia di condivisione delle risorse di calcolo e memoria.

Il termine *grid computing* venne coniato in un lungimirante libro di Ian Foster e Karl Kesselman *The Grid: blueprint for a new computing infrastructure* ("La Grid: il progetto di una nuova infrastruttura di calcolo"), pubblicato negli Stati Uniti nel 1999. La Grid veniva annunciata dagli autori come un naturale potenziamento della rete, una nuova applicazione in grado di rendere fruibile universalmente la potenza e le risorse di calcolo, fornite da centri sparsi in tutto il mondo, così come il *world wide web* aveva reso universale e libero l'accesso alle informazioni diffuse in rete. La metafora della "griglia" rifletteva l'idea che un utente possa utilizzare tutte le risorse di calcolo che gli occorrono con un semplice collegamento alla rete, senza sapere da dove vengono o come siano prodotte, come avviene, ad esempio, per la rete dell'energia elettrica.

E proprio alla fine dello stesso anno partiva il pionieristico progetto italiano Infn-Grid, che insieme al progetto europeo 2001 Datagrid e ai successivi progetti Egee (*Enabling Grids for E-science*) e infine il *World-wide Lhc Computing Grid (Wlcg)* coordinati dal Cern, hanno dimostrato la possibilità concreta di costruire una nuova infrastruttura dotata di queste capacità. Una Grid capace di fatto di fornire a tutto il mondo della ricerca un accesso uniforme e condiviso, non solo alle risorse di calcolo, ma anche ai grandi archivi di dati sparsi in vari paesi e collegati tramite Internet, che appaiono all'utente come una grande risorsa globale, dotata di un'unica interfaccia.

Le griglie rispondono a un bisogno che viene direttamente dal mondo della ricerca, come è accaduto

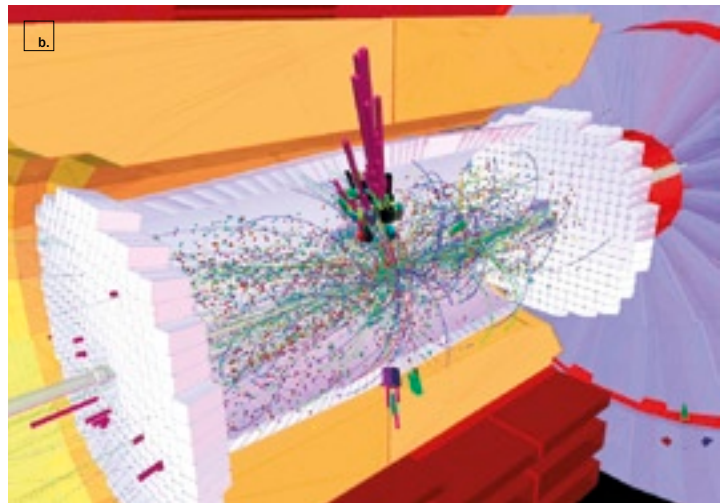
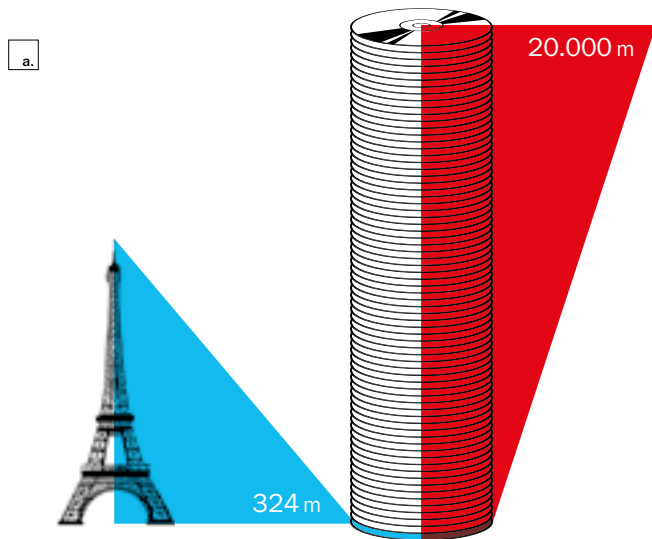
a suo tempo per il web. La condivisione di risorse di calcolo e archivi di dati facilitano infatti grandemente le attività delle comunità scientifiche e dei gruppi di ricerca differenti, che partecipano allo stesso progetto. Grazie inoltre alla semplificazione e uniformità delle regole di accesso, i ricercatori possono accedere in modo più semplice, agevole ed economicamente vantaggioso, a tutti gli archivi e alle risorse messe in comune, anche se non sono amministrare centralmente, ma appartengono a più proprietari completamente indipendenti. Questi standard di comunicazione e accesso sono inoltre generali e *open source*, ovvero possono essere acquisiti e usati liberamente

e gratuitamente, per garantire l'accesso alle risorse disponibili anche da parte degli utenti di altri settori scientifici. È necessario però attribuire ai membri delle varie comunità diritti di accesso e utilizzo ben definiti e definire e concordare politiche di sicurezza e di uso comuni.

Da questo punto di vista il modello Grid si è rivelato un grande successo. L'infrastruttura di Egee oggi in Europa permette ormai a più di 10.000 ricercatori delle più svariate discipline di utilizzare in modo uniforme per i propri calcoli e i propri dati più di 80.000 nodi di calcolo e 20 milioni di miliardi di byte (petabyte) di archivi, localizzati in circa 250 centri europei grandi

a. / b.

La Grid è stata progettata per immagazzinare e rendere accessibili i 15 milioni di Gigabyte di dati che saranno prodotti ogni anno da Lhc, pari a una torre di cd di 20.000 metri, circa 61 volte l'altezza della Torre Eiffel (324 m). A destra, la simulazione di uno dei milioni di eventi che saranno prodotti in Lhc: i prodotti di una collisione tra frasci di protoni registrati nel rivelatore Cms.



[as] box

Rapidi come nuvole

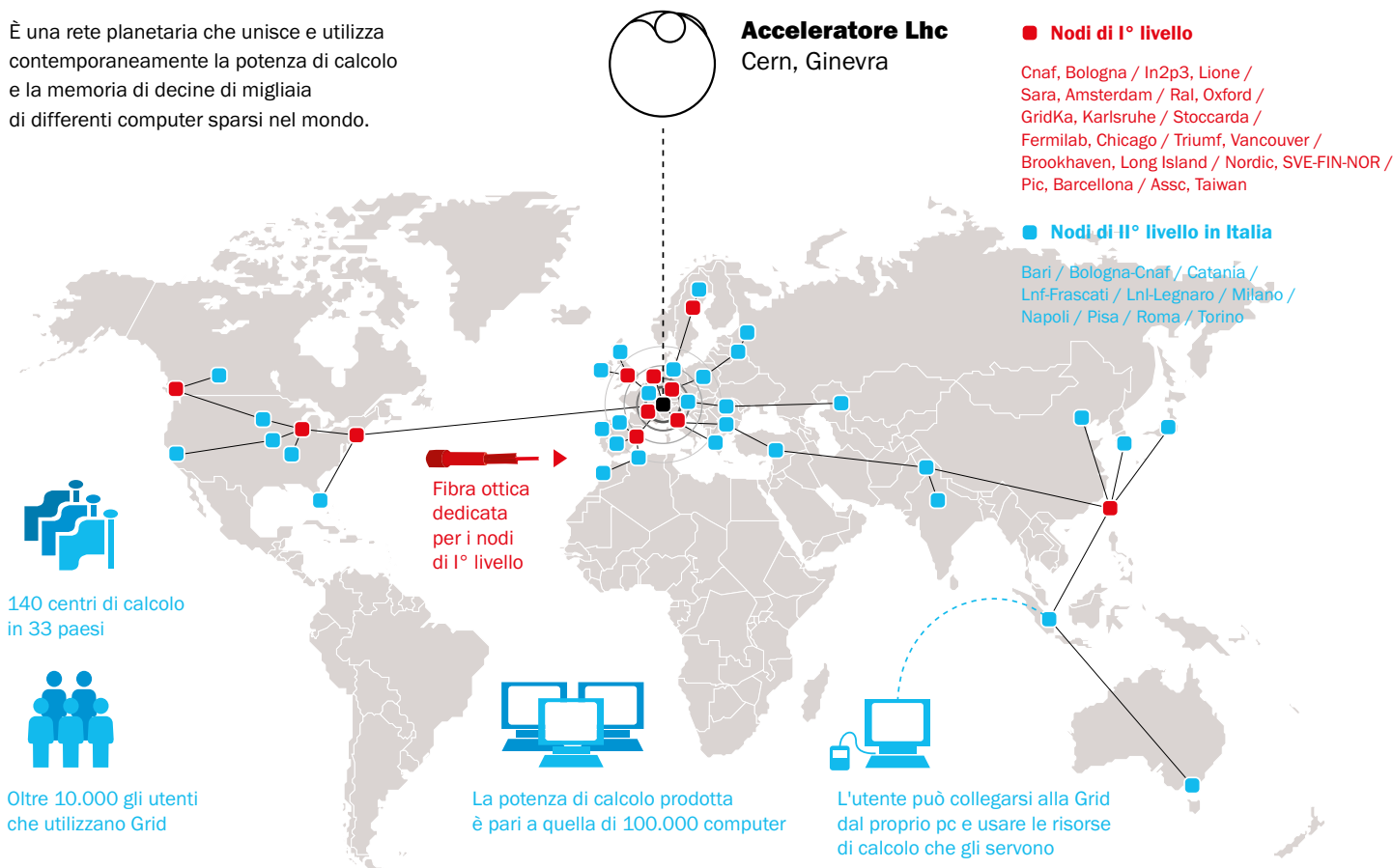
Nell'ultimo periodo una nuova applicazione proposta a livello commerciale dalle grandi compagnie Usa (ad esempio *Google* e *Amazon*), chiamata *Cloud Computing*, ha cominciato a offrire con successo agli utenti la possibilità di ospitare a pagamento le proprie applicazioni e i propri dati su un'unica risorsa di calcolo e memoria centrale, sulla base di protocolli privati e rinunciando a ogni tentativo di standard comuni di accesso da parte dei differenti fornitori. Il *Cloud Computing* offre un servizio che permette agli utenti di accedere ad ambienti virtuali creati secondo le loro necessità, ma senza avere nessun controllo sopra l'infrastruttura tecnologica che li supporta, nè garanzie di poter cambiare fornitore, senza dover imbarcarsi in un lavoro tedioso di adattamento. Le risorse e le interfacce sono amministrare centralmente da un unico proprietario, che agisce in completa autonomia con lo scopo di fidelizzare

il cliente. Un'analisi preliminare effettuata dal Cern e da Egee ha mostrato che questa offerta, oltre a non essere ancora economicamente vantaggiosa rispetto ai sistemi commerciali in uso nel mondo accademico, è oggi più adatta ad applicazioni semplici che alle complesse applicazioni scientifiche, che richiedono architetture di calcolo integrate con quelle dei grandi archivi di dati. Ha tuttavia riscosso un notevole successo per la facilità d'uso da parte degli utenti commerciali, che in questo modo hanno potuto eliminare i loro centri privati, quando questi non erano pienamente utilizzati o comunque non economicamente convenienti. In futuro, se ci sarà una convenienza economica, è possibile immaginare le risorse appartenenti a fornitori commerciali di *Cloud Computing* tra quelle condivise, grazie alla Grid, dalle comunità di ricerca.

[Mirco Mazzucato]

Che cos'è la Grid

È una rete planetaria che unisce e utilizza contemporaneamente la potenza di calcolo e la memoria di decine di migliaia di differenti computer sparsi nel mondo.



e piccoli, che prima erano utilizzabili solo a livello locale, con un aumento costante della potenza offerta e del numero di utilizzatori. La Grid non solo fornisce la "colla" tecnologica per raggiungere l'obiettivo di un uso efficiente di tutte le risorse disponibili, ma ne definisce anche le politiche necessarie sia per gli utilizzatori che per i fornitori di risorse.

La Grid dunque viene oggi utilizzata dagli scienziati per affrontare problemi cruciali ed estremamente complessi come studiare l'origine e l'evoluzione dell'Universo, costruire molecole in grado di debellare la malaria o combattere il cancro, prevedere l'effetto sulle nostre vite del riscaldamento globale, gestire situazioni di crisi ambientali o sanitarie.

Tuttavia, anche se il ruolo essenziale della standardizzazione per una diffusione generale delle griglie è sempre stato riconosciuto da

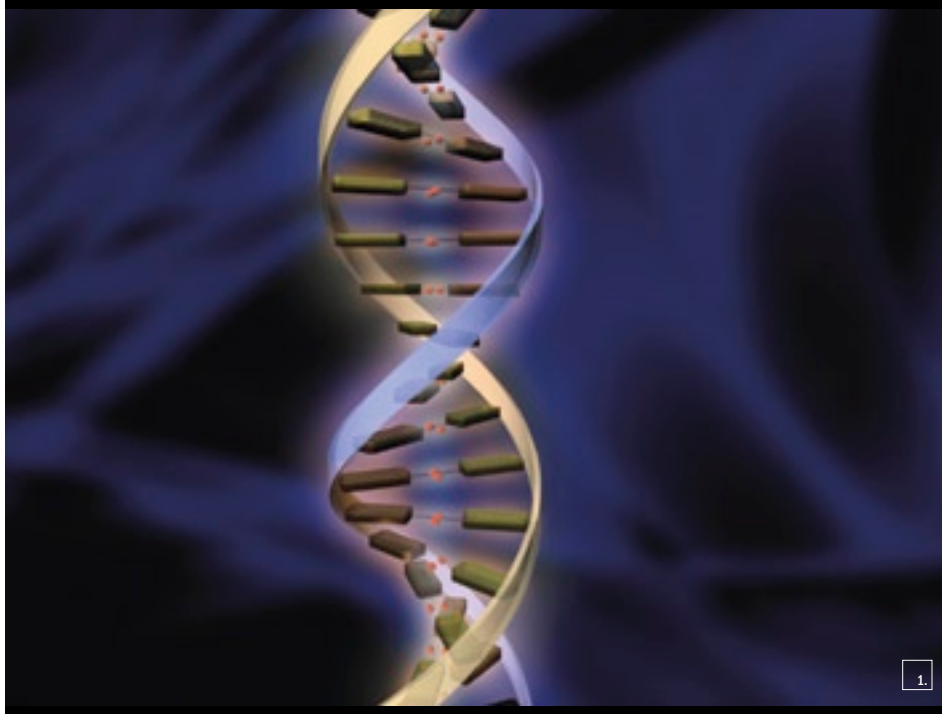
tutti, esigenze pratiche di una rapida risposta ai bisogni specifici di classi di utenti diversi hanno portato, come ai tempi della nascita di Internet, allo sviluppo di protocolli grid indipendenti che sono stati poi utilizzati nelle varie grid nazionali come, ad esempio, Egee in Europa, Osg negli USA, Naregi in Giappone. Questo ha generato difficoltà di comunicazione tra le differenti Grid, che solo recentemente hanno cominciato a essere affrontati sistematicamente e risolti. Ed è proprio il progetto *World-wide Lhc Computing Grid (Wlcg)* dei fisici che lavorano al Large Hadron Collider del Cern a essere il primo esempio di un'infrastruttura Grid mondiale, in cui i vari spezzoni a livello europeo, statunitense e asiatico, se pure dotati di servizi grid diversi, riescono a operare insieme grazie a uno sforzo d'integrazione delle varie interfacce e all'utilizzo di alcuni servizi comuni.

La Grid che serve alla medicina

I servizi grid sviluppati per la fisica delle alte energie consentono di costruire archivi condivisi ed elaborare dati, distribuiti a livello planetario in centri di calcolo diversi, esigenza che si ripresenta con molte analogie in medicina. La sempre maggiore importanza della prevenzione e quindi delle tecnologie di *imaging* digitale comporta infatti una produzione di grandi quantità di dati, distribuiti geograficamente, poiché provengono dalle numerose strutture presenti sul territorio.

Alcune applicazioni mediche di servizi grid esistono già, in particolare nei casi in cui siano necessarie molte risorse di calcolo e brevi tempi di risposta, impiegando in parallelo risorse di calcolo e memoria differenti; in altri casi esistono dei prototipi, ad esempio per la ricerca tramite algoritmi di anomalie in immagini

mediche (mammografie, tomografie del polmone). C'è tuttavia un punto cruciale che differenzia le applicazioni mediche da quelle di fisica delle particelle: la riservatezza dei dati e il controllo rigoroso degli accessi, che rende opportuna una Grid dedicata alla medicina, rigorosamente inaccessibile ad altre applicazioni. La tecnologia per garantire questa funzionalità esiste. All'orizzonte ci potrebbe quindi essere un sistema integrato con servizi grid e web per la gestione ed elaborazione di dati e immagini dei pazienti, per la pianificazione dell'accesso ai servizi sanitari e per la gestione personalizzata e l'accesso tramite Internet alle cartelle cliniche personali, come se si trattasse di una speciale casella di posta elettronica.
[Piergiorgio Cerello]



1.
La Grid oggi è già utilizzata per calcoli o simulazioni legati alla biologia molecolare o alla previsione dei cambiamenti climatici.

Biografia

Piergiorgio Cerello è ricercatore presso la sezione Infn di Torino. Collabora all'esperimento Alice di Lhc e coordina il progetto MAGIC-5 per lo sviluppo di algoritmi di elaborazione di immagini mediche.

Biografia

Mirco Mazzucato è direttore del Cnaf, il Centro Nazionale di Calcolo avanzato dell'Infn. Rappresenta l'Infn nel progetto World Lhc Computing Grid del Cern e coordina la ricerca per la costruzione dell'*Italian Grid Infrastructure* (Igi). È stato coordinatore italiano di alcuni esperimenti di fisica delle alte energie attuati al Cern.

Link sul web

<http://grid.infn.it>
<http://lcg.web.cern.ch/LCG>
<http://www.eu-egee.org>
<http://www.italiangrid.org>
<http://www.isgtw.org>