

## **Risorse per il calcolo in ENEA e alcune considerazioni sulla armonizzazione delle infrastrutture nazionali cloud per modellistica, simulazione e trattamento dati in ambito scientifico.**

G.Bracco, S.Migliori, R.Guadagni, A.Quintiliani, S.Podda, A.Funel, D.Abate, F.Ambrosino, T.Bastianelli, F.Beone, M.Chinnici, A.Colavincenzo, A.Cucurullo, P.Dangelo, M.Derosa, G.Furini, D.Giammattei, G.Giannini, S.Giusepponi, G.Guarnieri, A.Italiano, A.Mariano, G.Mencuccini, C.Mercuri, P.Ornelli, S.Pecoraro, A.Perozziello, A.Petricca, S.Pierattini, F.Poggi, G.Ponti, A.Rocchi, C.Scio, F.Simoni

### **Introduzione**

L'unità ICT [1] di ENEA, l' "Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile", in più di un decennio ha sviluppato la propria infrastruttura informatica dedicata al calcolo scientifico seguendo da una parte l'evoluzione dell'hardware messo a disposizione dal tumultuoso sviluppo tecnologico avvenuto e dall'altra parte implementando architetture software capaci di fornire all'utenza ambienti di calcolo per quanto possibile stabili ed omogenei, pur garantendo prestazioni allo stato dell'arte.

L'utenza dei sistemi ENEA è composta sia dai ricercatori interni all'istituzione che dai loro collaboratori esterni nell'ambito dei progetti di cui ENEA è partner. Esistono poi casi limitati ma significativi di utilizzo delle risorse di calcolo da parte dell'industria, nel quadro di accordi operanti nell'ambito delle attività di ricerca. I settori più significativi di utilizzo sono la modellistica e la simulazione in ambito della fluidodinamica, incluso il caso della combustione, la climatologia, inclusa la diffusione degli inquinanti, la chimica computazionale, il settore nucleare, fissione e fusione. Qualche attività anche nel campo della bioinformatica.

In tale contesto, tenuto conto della distribuzione geografica delle risorse di calcolo, dovuta anche all'organizzazione di ENEA in molti siti sul territorio nazionale, è stato utilizzato sin dall'inizio l'approccio della griglia computazionale che ha permesso di fornire all'utenza una visione unitaria delle risorse offerte privilegiando l'accesso in termini di servizi. In tal senso l'emergere recente del paradigma di cloud computing ha trovato in ENEAGRID [2] un terreno naturale per implementazioni sia specificatamente cloud con infrastrutture virtualizzate che in senso più lato in termini di focalizzazione sui servizi offerti all'utente finale.

Bisogna poi sottolineare che le attività descritte si sono accompagnate con la partecipazione a progetti nazionali ed internazionali come Datagrid, EGEE, BeinGrid, EGI, IGI e ad esperienze con svariati middleware, come globus, glite, unicorn. Da notare però che nella implementazione della propria infrastruttura software la scelta di ENEAGRID è stata quella di utilizzare prodotti maturi e multiplatforma, puntando nello sviluppo soprattutto all'integrazione tra i componenti di base al fine di fornire servizi mirati all'utenza.

Nelle due sezioni seguenti vengono rispettivamente (1) riportate sinteticamente alcune osservazioni sulle prospettive di armonizzazione delle risorse di calcolo di ENEA in ambito cloud nazionale e (2) descritte in qualche dettaglio tali risorse, inclusa la loro evoluzione a breve.

### **Sulla armonizzazione delle infrastrutture cloud nazionali**

Il quadro nazionale della organizzazione delle risorse dedicate al calcolo scientifico attraversa un momento di importanti trasformazioni. I finanziamenti PON 2002-2006 hanno portato negli anni passati alla creazione nell'Italia meridionale di nuove realtà focalizzate al calcolo scientifico, con i progetti CRESCO, COMETA, SCOPE e CYBERSAR e il nuovo piano 2007-2013 sta portando anch'esso alla implementazione di nuove risorse importanti. Da poco poi è iniziato il processo di unificazione in una unica istituzione, il CINECA, dei grandi consorzi di calcolo interuniversitari, CINECA, CASPUR e CILEA. A ciò si aggiunge il manifestarsi dei riflessi in Italia dell'evoluzione dei

progetti europei della linea EGEE/EGI, che ha portato prima alla creazione della JRU IGI, quindi a una attribuzione di fondi ministeriali specifici per IGI ad INFN. Infine l'attuarsi della modifica di statuto GARR che assegna a tale istituzione un ruolo specifico di sviluppo e fornitura di servizi nel campo dell'accesso alle risorse di calcolo e storage per la comunità scientifica. Questi ultimi temi sono quelli che più specificatamente possono essere affrontati nel quadro del paradigma delle cloud.

In tale contesto, a valle dell'esperienza maturata in ENEA e tenendo conto del confronto con quanto avviene anche in altri ambiti nazionali, come quelli europei, proponiamo i seguenti spunti di analisi, qui esposti in modo molto sintetico:

- L'attività di modellistica e simulazione richiede infrastrutture di calcolo e storage di tipologie diverse, in modo molto sintetico: (1) sistemi a larghissimo parallelismo, al limite delle tecnologie attuali; (2) sistemi di taglia media orientati alle applicazioni di media scalabilità; (3) sistemi a grande memoria condivisa; (4) sistemi per il calcolo massicciamente distribuito. L'ambito dell'approccio cloud esclude sostanzialmente la prima delle precedenti specificità ma comprende in modo vario le altre tre, che indicheremo nel seguito come servizi di calcolo avanzato.
- L'unificazione delle istituzioni più specificatamente orientate al calcolo ad alte prestazioni se da una parte può dar luogo ad una promettente economia di scala a livello degli investimenti nazionali nel campo, può d'altro canto risultare un fattore limitante degli effetti benefici della competizione tra diversi attori, soprattutto se si focalizza l'attenzione su sistemi di calcolo avanzato escludendo il terreno dei sistemi ad altissimo parallelismo, per i quali ogni nazione tende ad avere un numero molto limitato di siti di riferimento.
- Gli effetti benefici di un regime di competizione tra fornitori di infrastrutture di calcolo avanzato diventano anche elementi imprescindibili del quadro attuale del sistema di finanziamento della ricerca basato sulla selezione di progetti espressi a fronte di bandi le cui specificità spesso richiedono importanti risorse di calcolo per essere portate a buon fine. L'approccio cloud permette di fornire interfacce standard anche ad utenze esterne alla istituzione ospitante le risorse di calcolo facilitando quindi lo sviluppo di collaborazioni negli ambiti progettuali.
- L'evoluzione tecnologica dei sistemi di calcolo sta puntando fortemente sulla riduzione dei consumi energetici a parità di potenza di risorsa computazionale. Tale evoluzione in presenza di infrastrutture già esistenti nei siti delle varie istituzioni, almeno in termini potenza elettrica e capacità di raffreddamento, abilita la possibilità di ospitare nuove risorse di calcolo richieste da esigenze di progetto, ottimizzando l'utilizzo dei fondi allocati e offrendo nuove occasioni di collaborazioni progettuali per le quali la modalità cloud potrebbe presentare vantaggi all'utente finale.
- Storicamente le varie istituzioni nazionali che attualmente offrono servizi di calcolo scientifico avanzato hanno sviluppato soluzioni hardware e software che nascono da una combinazione tra le richieste specifiche della propria utenza e dall'ammontare delle risorse disponibili. Con un'attenta partecipazione a progetti è possibile sia mantenere aggiornate le risorse di calcolo interne che sviluppare soluzioni e modalità di accesso che si adattano alle esigenze specifiche delle comunità coinvolte. In tal senso sono state implementate e sono tuttora in evoluzione varie soluzioni cloud di caratteristiche diverse.

Dall'esame dei punti precedenti emerge la motivazione per una attività di coordinamento ed armonizzazione nell'ambito della fornitura di servizi e modalità di accesso di base a servizi di cloud che permetterebbe, pur mantenendo un livello ragionevole di competizione tra realtà diverse, di ottimizzare l'utilizzo delle risorse disponibili per il calcolo scientifico avanzato nazionale.

In tal senso l'estensione del mandato di GARR appare bene inserirsi in una prospettiva di sviluppo ottimizzato delle risorse di calcolo nazionali, sviluppo nel quale le esigenze di armonizzazione dovranno combinarsi con l'esigenza di flessibilità richiesta dal rapido evolvere degli strumenti messi a disposizione dall'evoluzione tecnologica del settore e con una ragionevole garanzia di autonomia delle varie istituzioni partecipanti.

In questo quadro ENEA può contribuire sia esprimendo le esigenze della sua specifica comunità di utenti che, come si è sottolineato in precedenza, copre un ampio spettro delle tipologie applicative sia partecipando come fornitore di servizi di calcolo al sistema del calcolo scientifico nazionale sia in termini di capacità elaborativa che di soluzioni tecnologiche.

### **L'infrastruttura del calcolo scientifico di ENEA e la sua evoluzione a breve**

ENEAGRID integra in una unica infrastruttura l'insieme delle risorse computazionali dedicate al calcolo scientifico di ENEA. ENEAGRID fornisce un ambiente di lavoro omogeneo e metodi di accessi unificati mettendo a disposizione ai suoi utenti l'accesso a diversi sistemi operativi e a varie architetture hardware (Linux X86\_64, AIX SP5, GPU) come anche a molti software, proprietari oppure open source. Come conseguenza la modalità di lavoro dei ricercatori ENEA e dei loro collaboratori è indipendente dalla posizione geografica della risorsa hardware/software richiesta. Questo approccio favorisce le attività collaborative e la condivisione dei risultati tra i vari utenti e facilita anche l'amministrazione dell'infrastruttura. I siti di ENEAGRID sono connessi dalla rete messa a disposizione da GARR, la rete della ricerca e della accademia italiana [3].

ENEAGRID supporta un ampio spettro di applicazioni nei vari campi della simulazione, modellazione, analisi di dati, visualizzazione anche tridimensionale ed è connessa anche ad apparati sperimentali (ad. es. microscopi elettronici per l'analisi dei materiali, tavole vibranti per lo studio dell'effetto dei sismi) tramite opportune interfacce coi sistemi di controllo ed acquisizione dei dati. In molti contesti è stato poi utilizzato il paradigma dei laboratori virtuali, organizzati per aree tematiche e centralizzati su siti web dedicati. Un laboratorio virtuale è uno strumento di ENEAGRID che permette agli utenti di una comunità specifica di accedere in modo organico all'insieme di dati ed applicazioni rilevanti per quel settore. Recentemente alcune delle risorse di ENEAGRID sono state dedicate ad esperienze di virtualizzazione e cloud computing in un approccio integrato con l'infrastruttura esistente per esempio in termini di autenticazione e condivisione dati.

Le risorse computazionali più importanti di ENEAGRID sono i cluster HPC CRESCO [4] basati sulla tecnologia x86\_64. Questi cluster sono ospitati in 5 siti ENEA e sono stati acquisiti sia con fondi interni dell'istituzione che con fondi dei progetti finanziati dal MIUR[5]. I sistemi attualmente in linea forniscono circa 6000 core con una potenza computazione aggregata di circa 50 Tflops. Lo spazio dati in linea consiste attualmente di circa 300 TB. Nell'ambito dei nuovi progetti è prevista una espansione dei sistemi attuali con l'acquisizione di un nuovo singolo cluster da circa 5000 core ed uno spazio dati di 600 TB, sulla scala temporale dei prossimi 6 mesi.

### Referenze

- [1] <http://www.utict.enea.it>
- [2] <http://www.eneagrid.enea.it>
- [3] <http://www.garr.it/>
- [4] <http://www.cresco.enea.it/>
- [5] <http://www.istruzione.it>