

L'architettura di ENEA-GRID, infrastruttura distribuita per la condivisione di risorse informatiche dedicate al calcolo scientifico

S. Migliori, G. Bracco, P. D'Angelo

ENEA

Servizio Centralizzato Informatica e Reti

Lungotevere Thaon di Revel,76

00196, Roma

E-mails: migliori@enea.it, bracco@frascati.enea.it, dangelo@frascati.enea.it

Abstract

Nel testo viene brevemente descritta l'architettura di ENEA-GRID, inquadrandola nel panorama attuale delle attività relative all'impiego di griglie computazionali nel campo del calcolo scientifico. Vengono quindi illustrati i suoi componenti tecnici principali evidenziando infine le iniziative in corso di interconnessione con altre architetture di GRID.

Introduzione

La disponibilità di sistemi di calcolo potenti, distribuiti su vaste aree geografiche ma connessi tra loro da reti ad alta velocità, ha portato in questi anni allo sviluppo del concetto di griglia computazionale.

Il concetto di griglia (GRID) scaturisce dal tentativo di fornire una visione unitaria ad un insieme eterogeneo di risorse informatiche sparse tra siti distanti appartenenti, nel caso più generale, ad istituzioni con finalità ed attività del tutto diverse.

La realizzazione concreta di una griglia computazionale richiede la definizione di un insieme di strumenti standard, capaci di realizzare in modo uniforme l'accesso alle risorse informatiche disponibili, sia in termini di sistemi di calcolo che di archiviazione di dati. Inoltre tali strumenti devono essere in grado di garantire la sicurezza di operazione della griglia e fornire la possibilità di monitorare in ogni istante il suo funzionamento.

L'approccio basato sul concetto di griglia computazionale ha permesso di delineare un quadro unitario per l'insieme dei sistemi di calcolo di alte prestazioni disponibili all'interno dell'ENEA, al fine di ottimizzare il loro utilizzo all'interno di un ente distribuito sul tutto il territorio nazionale.

L'attività dell'ENEA infatti si svolge in 12 centri situati in Italia settentrionale, centrale e meridionale. Le sedi maggiori sono dotate ognuna di un centro di calcolo che gestisce le risorse installate localmente e che coprono esigenze di calcolo seriale, di calcolo parallelo e richieste di grafica avanzata.

ENEA-GRID

Il sistema ENEA-GRID [1] permette di accedere all'insieme delle risorse disponibili, distribuite geograficamente, come ad un unico sistema virtuale, con una capacità integrata globale dell'ordine di 1 TFlop, distribuita su sistemi paralleli multipiattaforma (AIX/IBM SP, Solaris/SUN, Linux x86/IA64/Alpha/, IRIX/SGI, Mac OSX/Apple G5).

I componenti principali che permettono ad ENEA-GRID di fornire all'utente

questo unico sistema virtuale sono essenzialmente tre:

- una interfaccia grafica di accesso in java su portale Citrix Metaframe
- un gestore delle risorse, LSF (Load Sharing Facility)
- il file system distribuito AFS (Andrew File System) attualmente disponibile in ambito Open Source sotto il nome di OpenAFS.

In fig.1 l'architettura di ENEA-GRID è schematizzata in modo grafico.

Il livello più basso è quello del sistema operativo delle varie piattaforme che condividono dati ed applicazioni attraverso il livello immediatamente successivo, quello del file system distribuito AFS. La distribuzione del carico di lavoro viene gestita a livello di ognuno dei siti dallo strato successivo, quello di LSF, mentre la distribuzione di carico tra i vari siti viene controllata da LSF Multicluster. In tale ambiente vengono eseguiti codici utente e codici commerciali. I livelli più alti sono quelli delle interfaccia grafiche di accesso che sono basate sia sull'accesso diretto a server Citrix (ICA) che sull'accesso tramite Web. Recentemente è stata iniziata l'integrazione, allo stato attuale ancora parziale, anche di risorse operanti con sistemi Microsoft Windows. Per garantire l'affidabilità del sistema è in corso anche l'implementazione di servizi di monitoring che coprono l'insieme dei componenti descritti.

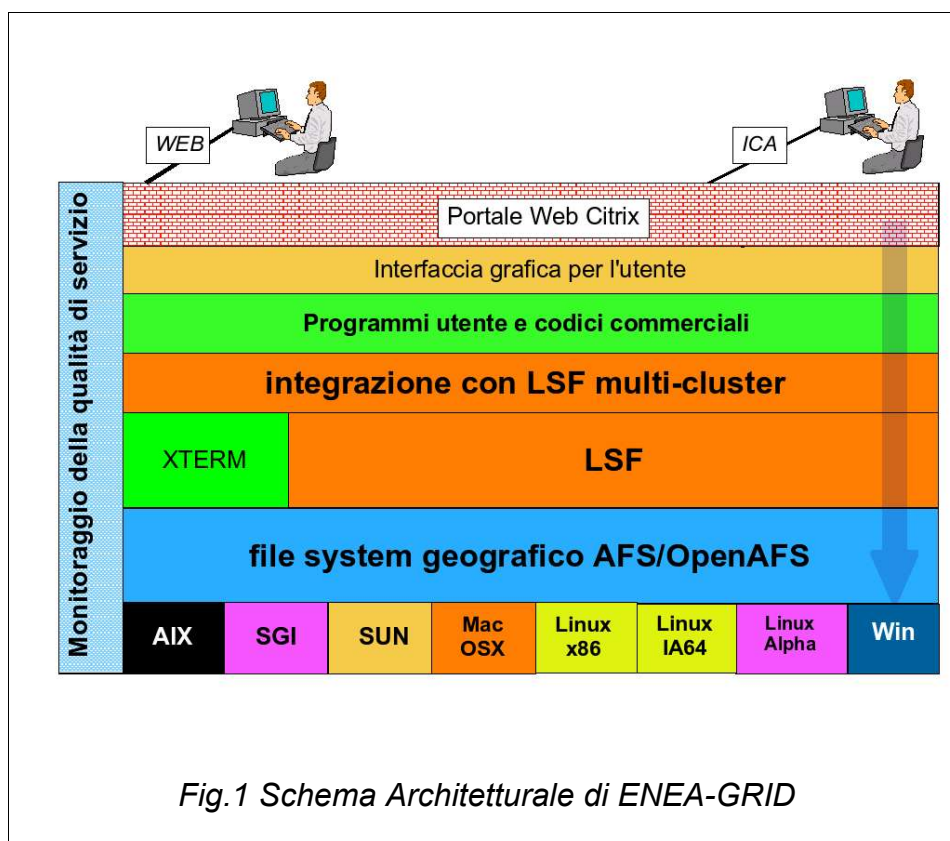


Fig.1 Schema Architeturale di ENEA-GRID

La scelta dei componenti architeturali di ENEA GRID si è rivolta a prodotti maturi, in parte proprietari e in parte in ambito OpenSource, al fine di ottimizzare le prestazioni ed i servizi offerti all'utenza dell'ente e di semplificare la gestione del sistema.

Lo sviluppo dell'infrastruttura è iniziato nel 1998 [3] e in essa sono state integrate successivamente le risorse di calcolo rese disponibili in ENEA-

INFO. Le risorse necessarie a realizzare l'intera infrastruttura sono in parte derivate dai fondi strutturali ENEA ed in parte dai finanziamenti MIUR per le aree ad obiettivo 1. Il finanziamento del progetto presentato dall'ENEA dal nome TELEGRID ha consentito la realizzazione dei cluster di Portici (Na) e Brindisi ed il potenziamento del cluster di Trisaia (MT).

Interoperabilità con altre architetture

In ambito internazionale sono in corso vari progetti per la realizzazione di griglie computazionali ed ENEA partecipa ad alcuni di essi, tra cui il progetto EGEE [2], che ha avuto inizio nel 2004. Il progetto EGEE costituisce la continuazione logica della attività del progetto DATAGRID, terminato all'inizio 2004, al quale ENEA aveva partecipato insieme a ESA (European Space Agency) con la realizzazione di un portale di accesso DATAGRID - ENEA GRID [3]. Tale portale aveva permesso l'utilizzo delle risorse informatiche ENEA per l'elaborazione dei dati satellitari prodotti dall'esperimento GOME, relativi alla misura del livello di Ozono. L'impegno di ENEA in EGEE è stato dichiarato nei termini di ~100 cpu al 20% del loro tempo, a partire dalla metà del 2005.

In ambito italiano ENEA partecipa all'attività del progetto FIRB Grid.it che nel quadro del Work Package 13, dedicato alla Chimica Computazionale, ha sviluppato una propria infrastruttura di calcolo in griglia. A tale infrastruttura ENEA contribuisce con la realizzazione di gateway basato su Globus, che permette l'accesso a risorse selezionate, disponibili in ENEA-GRID.

Integrazione con impianti sperimentali

Nell'implementazione del GRID-ENEA è stata rivolta sin dall'inizio una particolare attenzione alla sua integrazione con impianti sperimentali. Allo stato attuale risultano integrati tre impianti sperimentali:

- Sistema di acquisizione immagini ed elaborazione automatica relative ad un esperimento per la fusione termonucleare controllata installato presso il CR ENEA di Frascati (tokamak FTU),
- Un sequenziatore DNA per l'agricoltura presso il centro ENEA della Trisaia (MT);
- Un microscopio elettronico (TEM) da 300Kev installato presso il centro ENEA di Brindisi. Su questo sistema si stanno mettendo in atto tecniche di teleoperazione integrate in GRID.

In questo caso l'approccio seguito è stato quello di creare in prossimità degli impianti sperimentali delle opportune infrastrutture di integrazione che consentano l'immediata fruibilità dei dati prodotti dall'impianto a tutto il sistema di risorse hardware e software presenti in nel sistema.

Ringraziamenti

L'attività e lo sviluppo di ENEA-GRID sono stati resi possibili dalla collaborazione dell'insieme del Servizio Centralizzato Informatica e Reti di ENEA ed in particolare di S. Taglienti, R. Guadagni, S. Pecoraro, A. Perrozzello, D. Giammattei, A. Palumbo, M. Impara, S. Pierattini, M. Furini .

Un particolare ringraziamento va a A. Quintiliani che ha gestito tutta l'amministrazione verso MIUR per il progetto Telegrid.

Gli autori

Silvio Migliori, 1955, Laurea in Ingegneria Nucleare all'Università di Roma "La Sapienza", in ENEA dal 1983 dopo esperienza in Italcable, ha lavorato inizialmente nel Dipartimento Fusione Nucleare Controllata nella progettazione e nella realizzazione di impianti dedicati alla ricerca sperimentale. Fa parte dal 1995 del Servizio Informatica e Reti nel quale dal 2001 è il Vice Direttore per il Calcolo Scientifico. I suoi interessi correnti sono orientati ai sistemi avanzati di infrastrutture informatiche innovative (griglie computazionali, ambienti virtuali)

Giovanni Bracco, 1954, Laurea in Fisica all'Università di Pavia, in ENEA dal 1981, ha lavorato fino al 2003 nel Dipartimento Fusione Nucleare Controllata partecipando alla realizzazione di apparati diagnostici e alla attività di analisi dati con particolare interesse allo sviluppo di sistemi informatici dedicati alla attività sperimentale. Dal 2003 nel Servizio Informatico e Reti si occupa di calcolo scientifico distribuito.

Pietro D'Angelo, 1955, Diploma Tecnico Industriale per Elettronici, in ENEA dal 1976 come amministratore dei sistemi di calcolo. Dal 1998 segue tutti gli aspetti implementativi, di sviluppo e manutenzione del sistema GRID ENEA.

Referenze

- [1] S. Migliori, G. Bracco, R. Guadagni, S. Taglienti, "ENEA-GRID: a production quality grid infrastructure", Incontro GARR "Armonizzazione delle strutture di rete ed delle griglie computazionali", CRUI, Roma, 15/7/2004, http://www.garr.it/incontro_griglie.htm
- [2] <http://public.eu-egee.org>; il contributo ENEA: G. Bracco, S. Migliori "Integrazione di ENEA GRID in EGEE", Third INFN-GRID Workshop, Catania, 15-17 ottobre 2003 <http://alipc1.ct.infn.it/workshop-infngrid>
- [3] S. Migliori, et al. "ENEA Computing Factory", Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications, PDPTA 1999, June 28 July 1, 1999, Las Vegas, Nevada, USA. Vol.6, 3037-3040