

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 INFO [Servizio Centralizzato Informatica e Reti], Roma
migliori@enea.it, bracco@frascati.enea.it

Contenuto

- Che cosa è ENEA-GRID
- Il paradigma di GRID & ENEA-GRID
- Integrazione con altre infrastrutture di GRID

Ente per le Nuove Tecnologie l'Energia e l'Ambiente

12 centri di ricerca

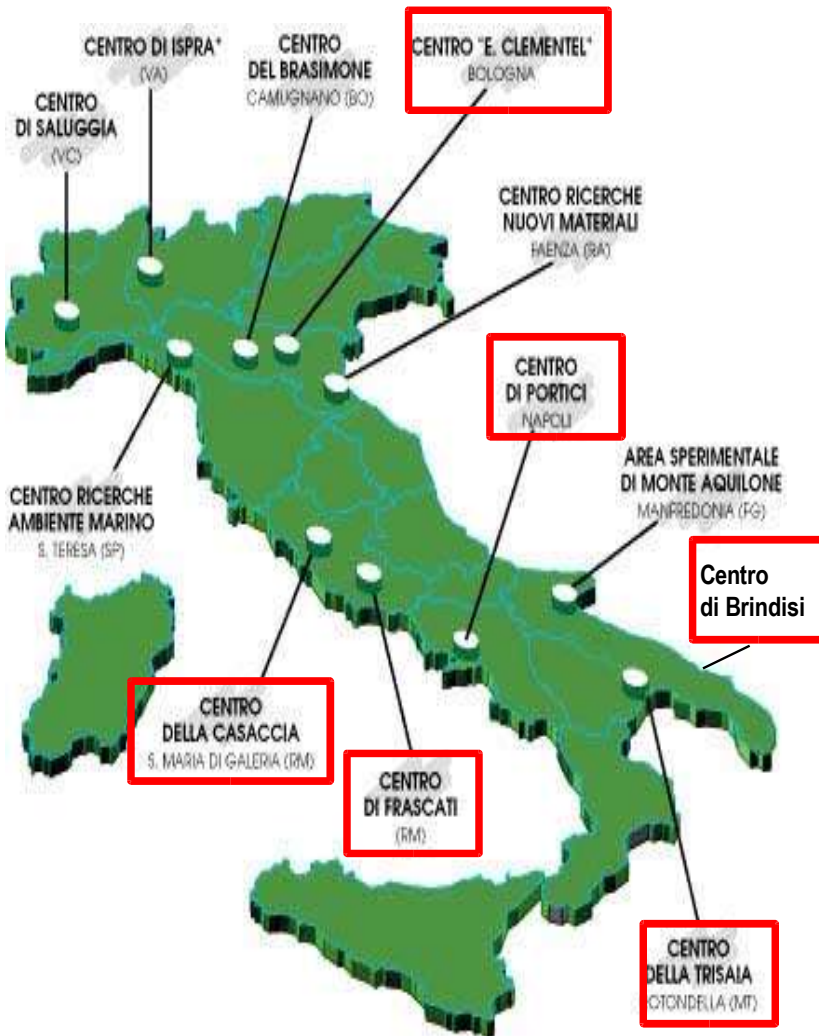
ENEA-INFO: il servizio Centralizzato
di Informatica e Reti

6 Centri di Calcolo

Casaccia, Frascati, Bologna, Trisaia,
Portici, Brindisi

Risorse Multiplatforma per il
calcolo seriale/parallelo e per il post
processamento grafico.

Risorse di calcolo dipartimentali ed
individuali



ENEA

Ente per le Nuove Tecnologie l'Energia e l'Ambiente

12 centri di ricerca

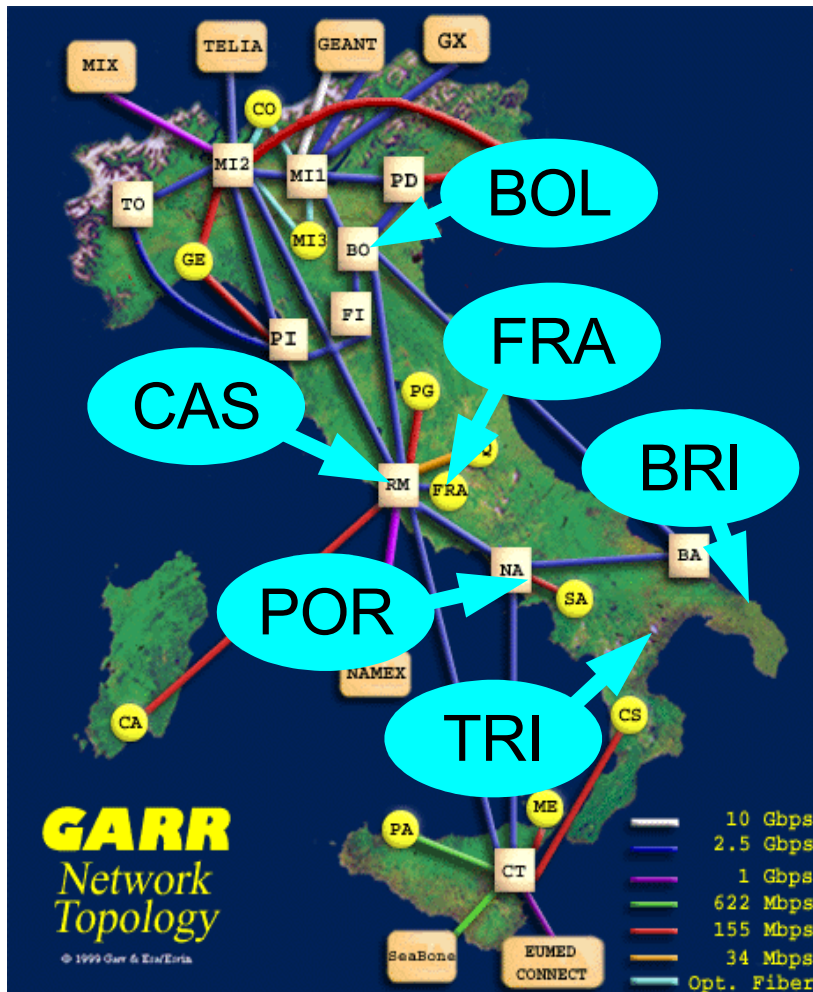
ENEA-INFO: il servizio Centralizzato
di Informatica e Reti

6 Centri di Calcolo

Casaccia, Frascati, Bologna, Trisaia,
Portici, Brindisi

Risorse Multiplatforma per il
calcolo seriale/parallelo e per il post
processamento grafico.

Risorse di calcolo dipartimentali ed
individuali



ENEA GRID: perchè

ENEA GRID è un sistema integrato capace di rispondere ai seguenti obiettivi:

- mettere a disposizione un **sistema di produzione** capace di offrire i **servizi necessari al calcolo scientifico**
- integrare l'insieme di risorse informatiche di ENEA-INFO, distribuite su WAN
- fornire un **ambiente di lavoro unificato** e **metodi di accesso omogenei** per tutti i ricercatori dell'ENEA indipendentemente dalla loro sede di lavoro.
- mettere a disposizione strumenti capaci di permettere l'**integrazione** delle **risorse informatiche dipartimentali** ed **individuali** e di fornire supporto alle **attività sperimentali**

I servizi di ENEA GRID

Servizi per il calcolo scientifico:

- **Sistemi paralleli/seriali multiplatforma**
 - AIX, Linux (Alpha,x86,x86_64,IA64), IRIX, Mac OS X, Solaris
- **Risorse 3D e il post-processamento grafico**
 - SGI + 3D immersivo
- **Risorse Software:**
 - Codici commerciali: Fluent, Gambit, Abacus, Catia, Ansys
 - Codici di “ricerca”: mcnp/x, eranos, fluka,....
 - Ambienti di sviluppo & analisi: IDL, Matlab, Mathematica, SAS
 - Applicazioni Windows

Risorse di Calcolo in ENEA-GRID

| OS | #cpu | Gflops | |
|-----------------|------|--------|---|
| AIX | 188 | ~1000 | Frascati(160),Bologna(8), Portici(18), Brindisi (2) |
| Linux x86 32/64 | 96 | 200 | Frascati(28), Casaccia(54), Portici(4), Trisaia(4), Brindisi(4) |
| Linux Alpha | 30 | 40 | Casaccia |
| IRIX | 27 | 40 | Frascati(8), Casaccia(4), Portici(1), Trisaia(8), Brindisi(1), Bologna(5) |
| Solaris | 8 | 10 | Trisaia(4), Casaccia(2), Bologna(?) |
| Windows | 18 | 50 | Frascati(6), Portici(4),Trisaia(4), Brindisi(4) |
| Mac OS X | 27 | 100 | Frascati(1), Trisaia(13) |

Le risorse di Portici, Brindisi & Trisaia sono state aumentate nel quadro del progetto
PON **TELEGRID**

L'architettura di ENEA GRID

La scelta di **componenti maturi** come garanzia di affidabilità e semplicità di gestione:

File system distribuito: **AFS/OpenAFS**

Gestore delle risorse: **LSF Multicluster**

Interfaccia grafica: **Java and Citrix Technologies**

Sistema di monitoring: **Patrol**

Server di licenze

Strumenti per l'integrazione delle risorse dipartimentali ed individuali

File sistem distribuito: AFS per software e dati

Condivisione del parco licenze

AFS: cella enea.it

AFS (Andrew File System), sviluppato presso Carnegie-Mellon University, commercializzato da Transarc & IBM, Open Source dal 2000, www.openafs.org.

Una architettura Client/Server matura [scalabile, multiplatforma].

Ottimizzato per funzionamento su WAN: replicazione dei dati su server, cache trasparente all'utente sul client.

Sicurezza: Kerberos, gestione nativa dell'autorizzazione.

AMBIENTE UNIFORME per l'utente

La HOME directory è la stessa su tutte le piattaforme UNIX/Linux ed essa è accessibile anche da Windows.

DISTRIBUZIONE UNIFORME del software anche multiplatforma

Semplificazione della gestione/distribuzione del software su WAN.

LSF

Load Sharing Facility con l'estensione MultiCluster
[www.platform.com]

Ambiente maturo e multiplatforma

Alta scalabilità (ENEA GRID ~300 cpu)

WAN: multicluster.

Certificato dai vendor principali (IBM, SGI,..) e integrato con il software commerciale.

Integrazione con AFS.

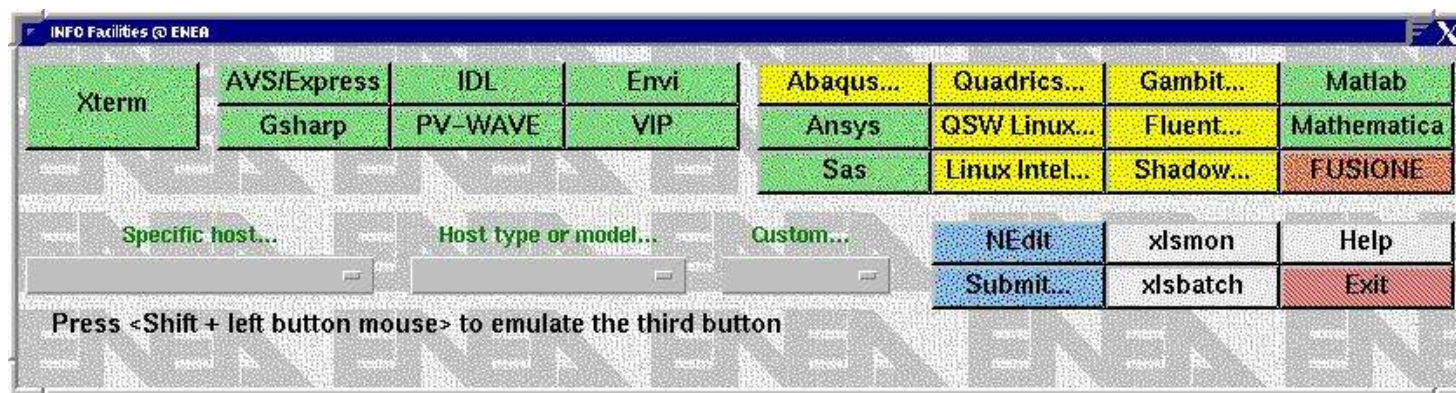
Flessibile: ad. es. gestione delle licenze.

ENEA GRID 6 LSF clusters:

Frascati, Bologna, Casaccia, Trisaia, Portici, Brindisi

Interfaccia utente

Accesso utente unificato tramite Citrix Presentation Center [ex Metaframe], java e script di shell:

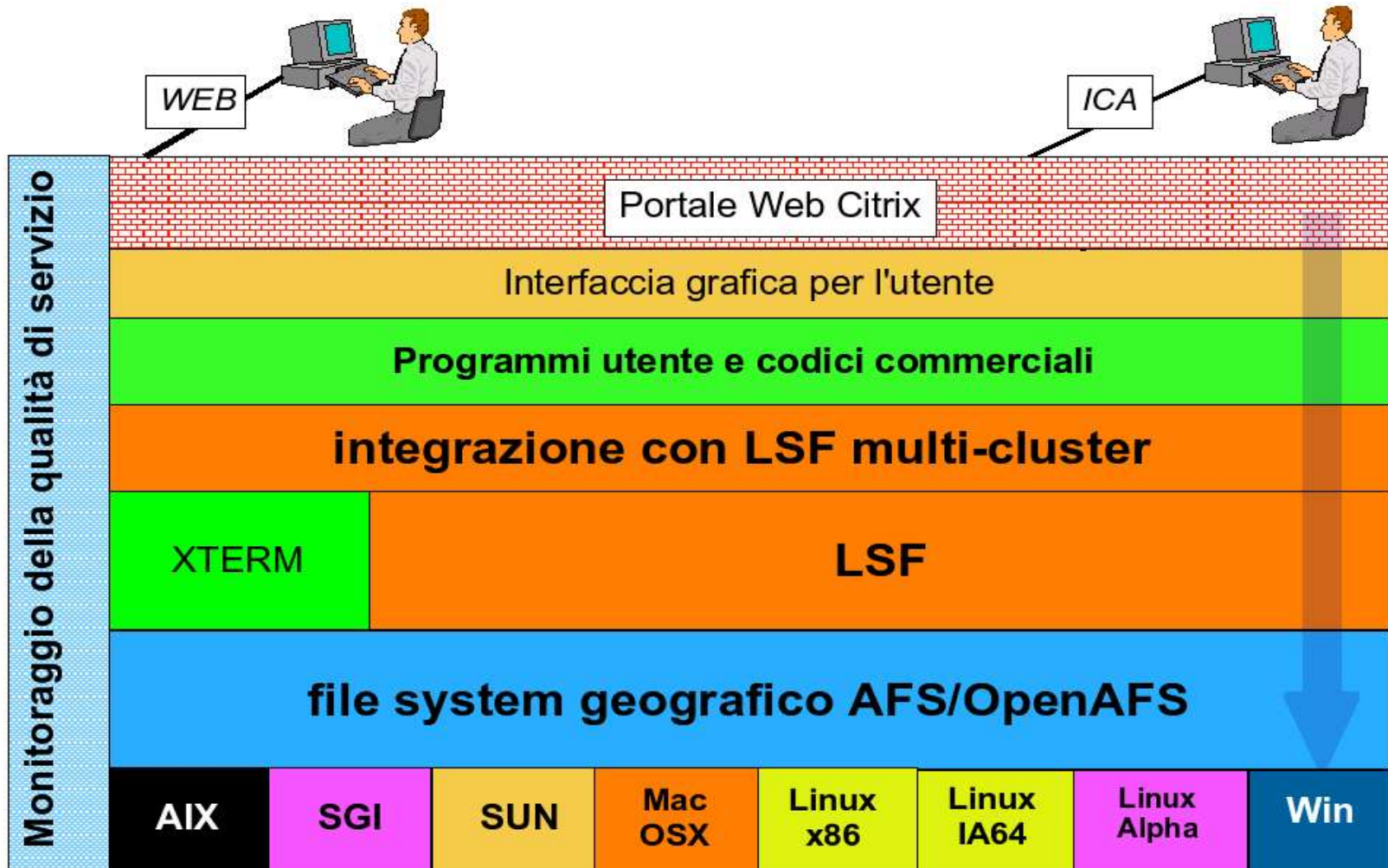


Prestazioni ottimizzate su WAN e metodi di comunicazione sicuri

Accesso gestito da 6 server Citrix UNIX [Bologna, Casaccia, Frascati, Trisaia, Portici, Brindisi]

4 server Citrix Windows forniscono accesso alle applicazioni **Windows** tramite Web(Frascati, Portici, Trisaia, Brindisi).

Architettura di ENEA GRID



ENEA GRID & il paradigma di GRID (1)

“GRID concept: coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations (VO)” Foster & Kesselman, The Anatomy of the Grid, 2001

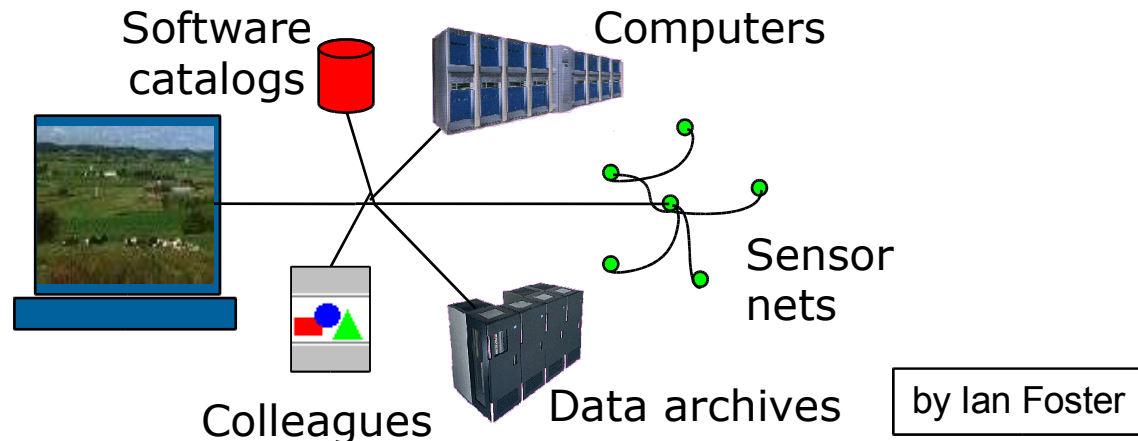
Il concetto di **VO concept** non può essere sviluppato a fondo in una singola organizzazione ma la complessità dell'ENEA costringe a ricercare strumenti con funzioni tipiche da VO (ad es. gruppi AFS, risorse LSF).

ENEA-GRID fornisce comunque la **condivisione coordinata delle risorse**, ottimizzando l'utilizzo di quanto messo a disposizione da ENEA INFO e promuovendo l'integrazione delle altre risorse interne all'ente.

Le funzionalità garantite dai componenti di ENEA GRID sono quelle tipiche del mondo GRID ed alcune ne costituiscono delle estensioni [ad es. gestione licenze]

ENEA GRID & il paradigma di GRID (2)

Quindi ENEA GRID fornisce una possibile risposta al “GRID Problem” : “unique authentication, authorization, resource access and resource discovery”, Foster & Kesselman 2001



La scelta di **componenti maturi** (LSF & Citrix - proprietari, AFS – progetto opensource dal 2000) riduce la generalità di ENEA GRID come modello di architettura di GRID ma ne aumenta l' **affidabilità** e la **facilità di gestione**.

L'esperienza di ENEA GRID

ENEA GRID è in produzione dal 1998

- Attualmente ~650 utenti registrati e ~2 TB di software e dati
- Le nuove risorse vengono sistematicamente integrate

L'utilizzo da parte dell'**utente** di modalità tipiche di GRID richiede un cambiamento di approccio all'accesso alle risorse di calcolo e l'accettazione dipende fortemente dalla **qualità del servizio**:

- “Valore” delle risorse messe a disposizione
- Efficienza della rete
- Monitoraggio del funzionamento complessivo
- Supporto all'utente per l'utilizzo dell'ambiente

La **gestione** di un sistema distribuito ma integrato in una organizzazione complessa come l'ENEA costituisce una sfida organizzativa, tecnica e culturale.

Integrazione con altri progetti di GRID

L'obiettivo della partecipazione di ENEA INFO in altri progetti di GRID è concentrato sulla **interoperabilità**.

Ogni progetto di GRID è caratterizzato da una propria architettura/middleware e l'accesso alle risorse deve risultare compatibile con l'infrastruttura interna:

==> realizzazione di gateway

Progetti:

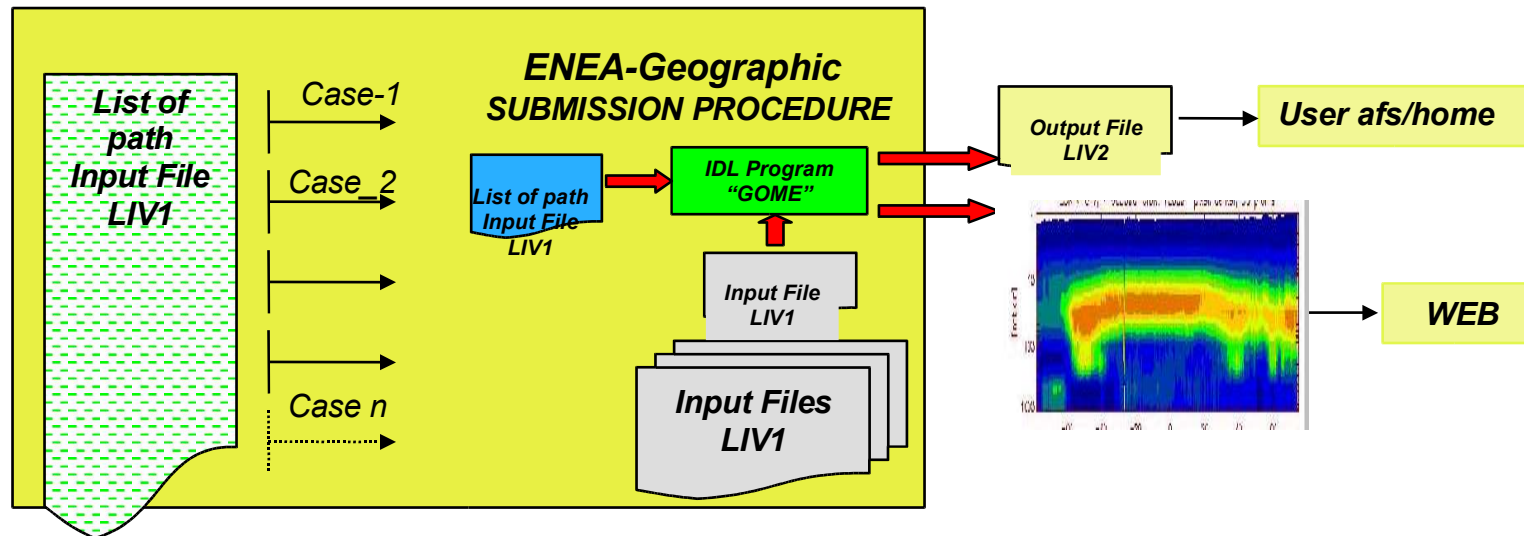
- Datagrid
- FIRB/Grid.it : WP13 Chimica Computazionale
- EGEE

ENEA GRID e il progetto DATAGRID

ENEA ha partecipato (unfunded) in DATAGRID (finito 2004) collaborando con ESRIN [ESA] Frascati dove è stato realizzato un gateway DATAGRID ==> ENEA-GRID con funzionalità limitate [sottomissione, pubblicazione dei risultati]

Test con l'analisi dei dati satellitari di GOME (livelli di Ozono)

[IDL/multicase/multiplatforma]



ENEA-GRID e il WP13 GRID.it/Chemgrid

Il progetto FIRB GRID.IT nel suo **WP13, Chimica Computazionale**, comprende un gruppo di ricercatori di ENEA Casaccia. WP13 ha realizzato una piccola GRID sperimentale basata su Globus 3.2 che comprende siti a PG, MI, BO, PD, BA e ENEA Frascati.

Il contributo ENEA: un **gateway** [chemgrid.frascati.enea.it] che permette l'accesso a 16 nodi Linux condivisi con ENEA-GRID.

I componenti installati:

Server: gatekeeper [compatibile con AFS ed LSF; MPICH-G2]

gsiftp [compatibile con AFS]

gssklogd [traduce certificato X509 in token kerberos AFS]

Client: globus 3.2 installato in AFS ed utilizzabile da ogni server

Linux di ENEA-GRID

ENEA-GRID e il progetto EGEE

EGEE (Enabling GRID for e-science in Europe) è un progetto del 6° programma quadro europeo che ha la missione di fornire **servizi GRID come sistema di produzione** [www.eu-egee.org]; inizio Aprile 2004.

ENEA è una dei partner [funded] all'interno della partecipazione italiana che è coordinata da INFN.

Impegno di ENEA: ~100 cpu (20%), seconda metà 2005.

Attualmente il middleware di EGEE è in corso di installazione a Frascati 4 macchine di servizio e 16 nodi di calcolo [Linux].

Si prevede l'accesso anche a risorse AIX tramite un gateway Linux

Ringraziamenti

L'attività di ENEA GRID è resa possibile da molti attori

in **ENEA**:

S. Taglienti, R. Guadagni, F. Simoni, A. Perozziello, A. De Gaetano, S. Pecoraro, D. Giammattei, M. De Rosa, M. Caiazzo, A. Palumbo, G. Elmo, S. Pierattini, M. Impara, G. Furini, C. Zini, V. Guastella, A. Quintiliani,...

da **Ditte** supporto software:

Nice, SysMan, Esse3Esse, CRIAI, BMC, Infordata,...

che hanno contribuito con l'attività di:

A. Secco, A. Falzone, C. Sciò, M. Fina, V. Pirone, A. Borelli, A. Brancaccio, L. Murianni,...