

Introduzione ai Sistemi di Grid e Cloud Computing

Obiettivi

- L'obiettivo di questa seconda parte del corso è lo studio dei sistemi di Cloud computing.
- Vengono presentati alcune tecnologie molto innovative dell'informatica distribuita che hanno punti di contatto molto importanti:
 - Grid computing,
 - Cloud computing.

Introduzione ai sistemi **Grid e Cloud** Definizioni

- La architettura delle Grid
- Il **Globus Toolkit** e altri sistemi (Condor, Unicore, gLite)
- Cloud Computing
- Service models dei Cloud

Griglie Computazionali

- **Obiettivi:**
- Costruire una rete di calcolo accessibile come la rete elettrica.
- Usare computer disponibili per costruire macchine virtuali distribuite ad alte prestazioni.
- Fornire accesso a servizi di elaborazione disponibili anywhere e anytime.
- Supportare la cooperazione di organizzazioni virtuali.

Griglie Computazionali

Forniscono accesso persistente e coordinato ad un insieme di risorse di calcolo connesse in rete



Applicazioni

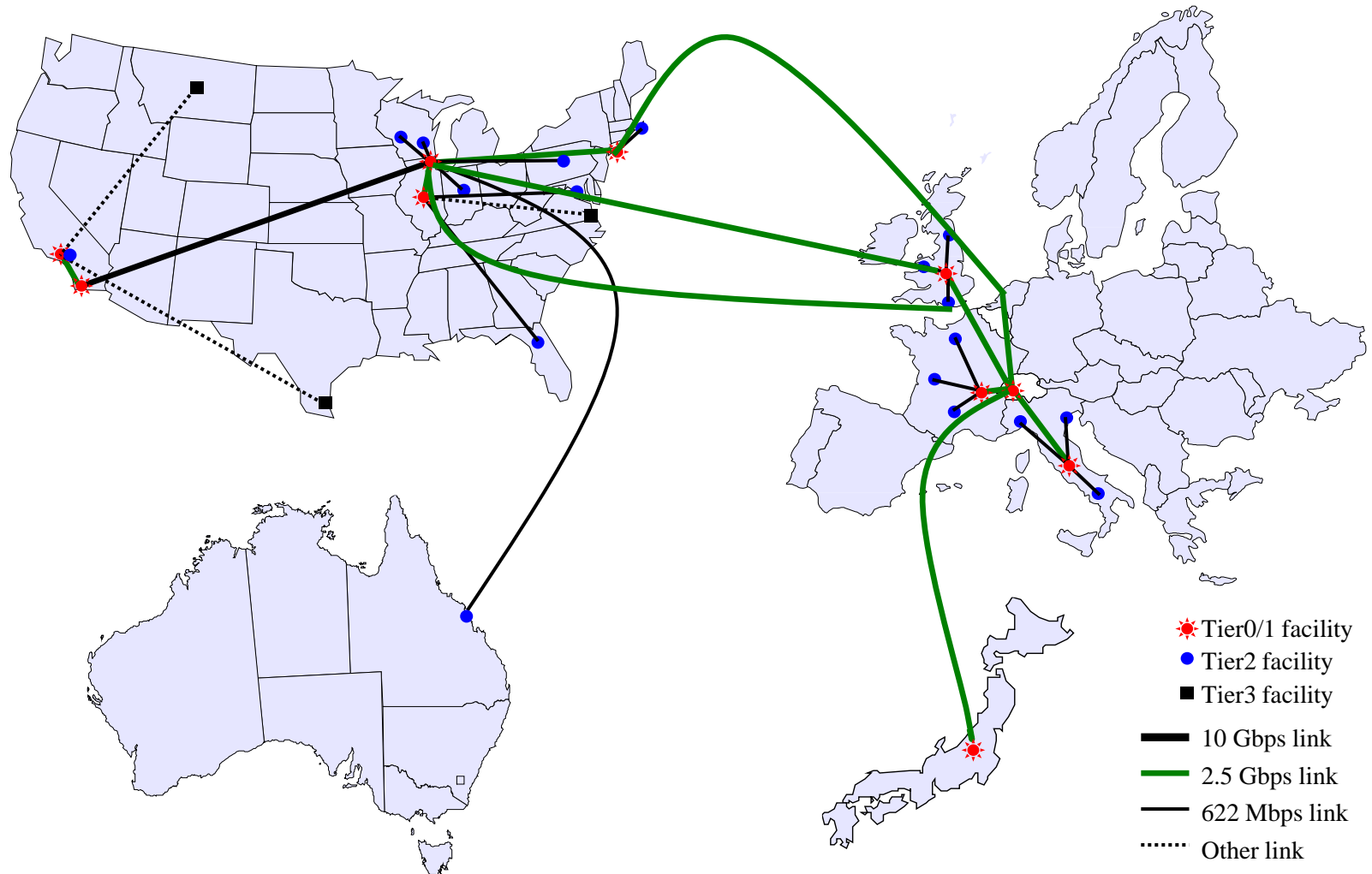
Supercalcolo distribuito
Gestione di grandi Database
Calcolo collaborativo



Organizzazioni Virtuali

Politiche di condivisione
Meccanismi di sicurezza
Natura dinamica

Griglie Computazionali : Un esempio



International Virtual Data Grid Laboratory

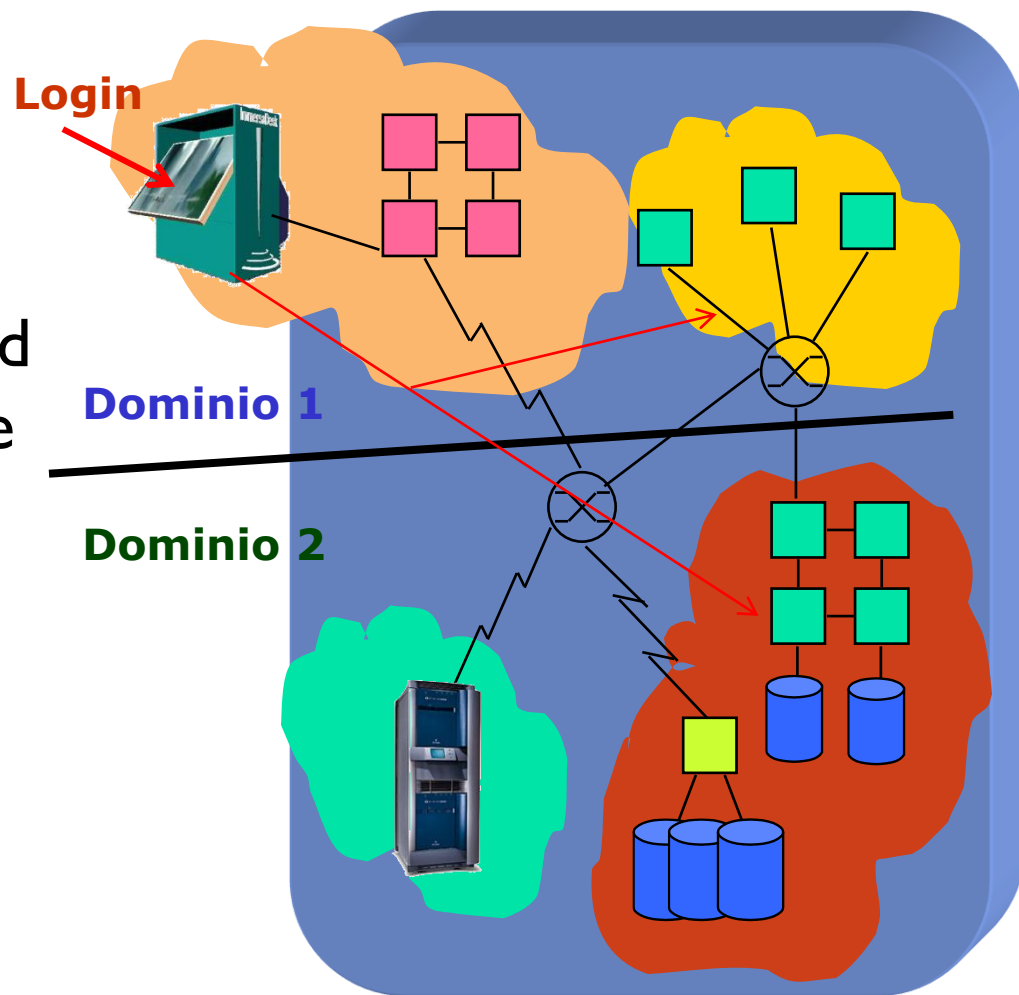
Griglie Computazionali

Ambienti che forniscono un accesso **affidabile, consistente, esteso e poco costoso** a risorse di calcolo ed informative geograficamente distribuite ed appartenenti ad organizzazioni diverse

- **affidabile**: garanzie di prestazioni predicibili e prolungate sui componenti del sistema
- **consistente**: interfacce uniformi ad un'ampia varietà di risorse e servizi standard
- **esteso**: possibilità di usare le risorse del sistema da qualsiasi punto di accesso
- **poco costoso**: accesso mediante strumenti hardware e software standard

Griglie Computazionali: Funzionalità

- Singola autenticazione
- Individuazione delle risorse
- Negoziazione dell'uso ed acquisizione delle risorse
- Esecuzione e gestione della computazione
- Accesso a dati remoti
- Analisi collaborativa dei risultati



GridComputing: Aree di Utilizzo

Distributed Supercomputing: risorse di calcolo aggregate per applicazioni ad alta complessità computazionale

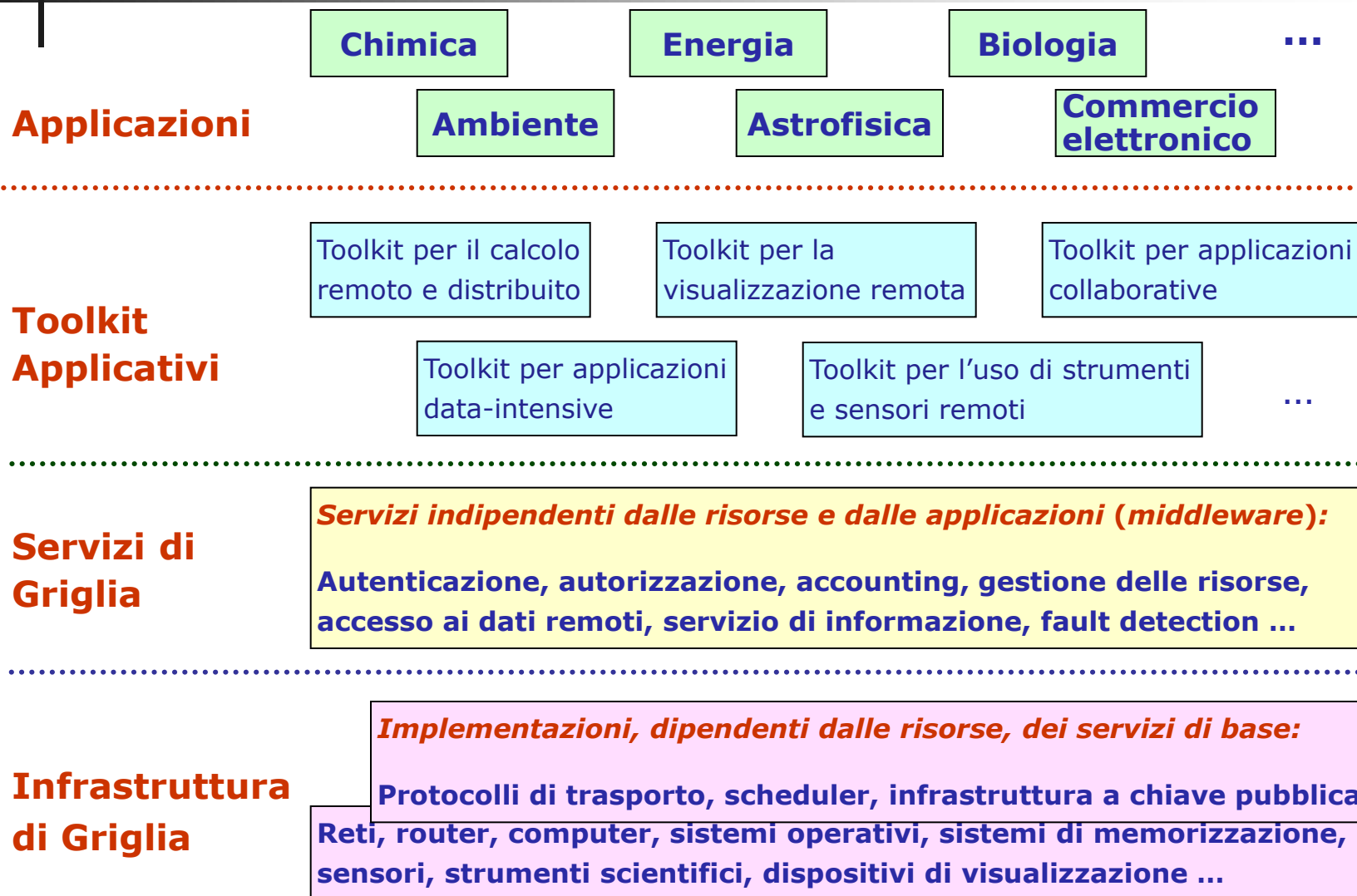
High-Throughput Computing: uso dei cicli di processori inutilizzati per ottenere lavoro utile

On-Demand Computing: soddisfare le necessità di breve termine di risorse di cui non è conveniente disporre localmente

Data-Intensive Computing: sintetizzare nuova informazione da dati mantenuti in database geograficamente distribuiti

Collaborative Computing: abilitazione ed intensificazione delle interazioni umane

Architettura di Griglia a Livelli



Ambienti per il Grid Computing

Condor:

- ambiente per l'high-throughput computing
- individuazione delle risorse inattive e migrazione dei job

Legion:

- singola macchina virtuale
- accesso trasparente alle risorse remote

Globus Toolkit:

- insieme di servizi e tool per il grid computing
- servizi globali costruiti su servizi locali
- da GT2 a GT4: dagli script ai grid services

Ambienti per il Grid Computing

gLite:

- ambiente per applicazioni scientifiche
- compute-intensive e high data transfer

UNICORE:

- ambiente di high performance computing
- esecuzione trasparente di job

Altri:

- Sun Grid Engine
- Grid Portals
- ...

Un esempio: Globus Toolkit 2

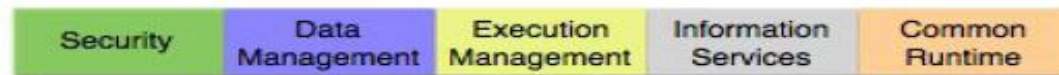
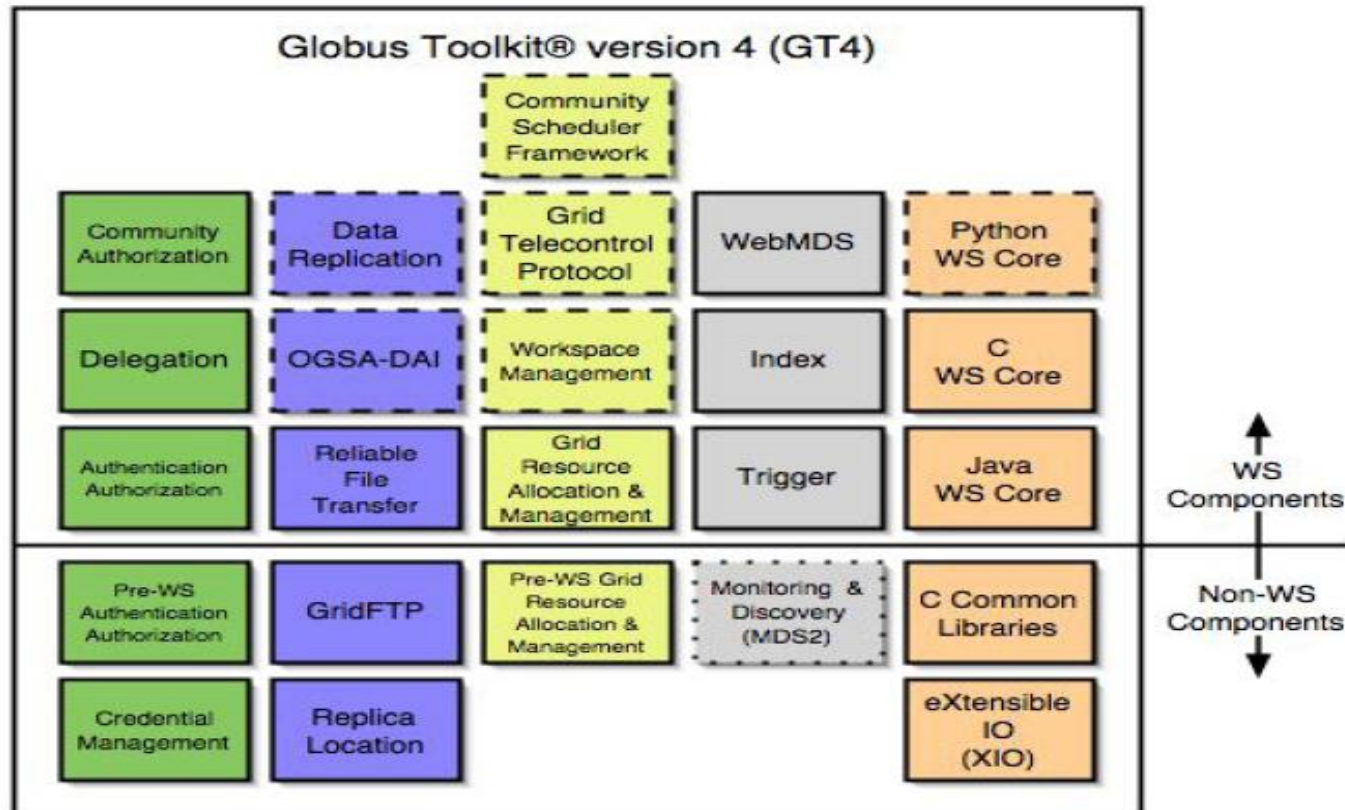
Approccio “**bag of services**”: un progetto modulare nel quale componenti distinti forniscono servizi per la gestione delle risorse, la sicurezza, l’informazione, ecc.




- **Globus Resource Allocation Manager (GRAM)**: allocazione delle risorse e creazione dei processi
- **Global Access to Secondary Storage (GASS)**: movimentazione ed accesso dei dati remoti
- **Grid Security Infrastructure (GSI)**: autenticazione con supporto per il controllo locale sui diritti di accesso
- **Grid Information Service (GIS)**: informazioni sullo stato dell’infrastruttura della griglia
- Comunicazione (**I/O, Nexus**)
- Monitoraggio dei componenti e rilevamento dei guasti (**HBM**).

Un esempio: Globus Toolkit 4

- Tutti i servizi e i client di GT2
- Basato sul modello WSRF (Web Services Resource Framework)
- Completa implementazione Java
- Services Globus “proprietary” costruiti su WSRF
 - Managed Jobs (simile al GT2 GRAM)
 - Reliable File Transfer (RFT)
 - Index Services (simile al GT2 GIS)

Architettura Globus Toolkit 4



-  Core GT Component: public interfaces frozen between incremental releases; best effort support
-  Contribution/Tech Preview: public interfaces may change between incremental releases
-  Deprecated Component: not supported; will be dropped in a future release

Gestione della Esecuzione di Job

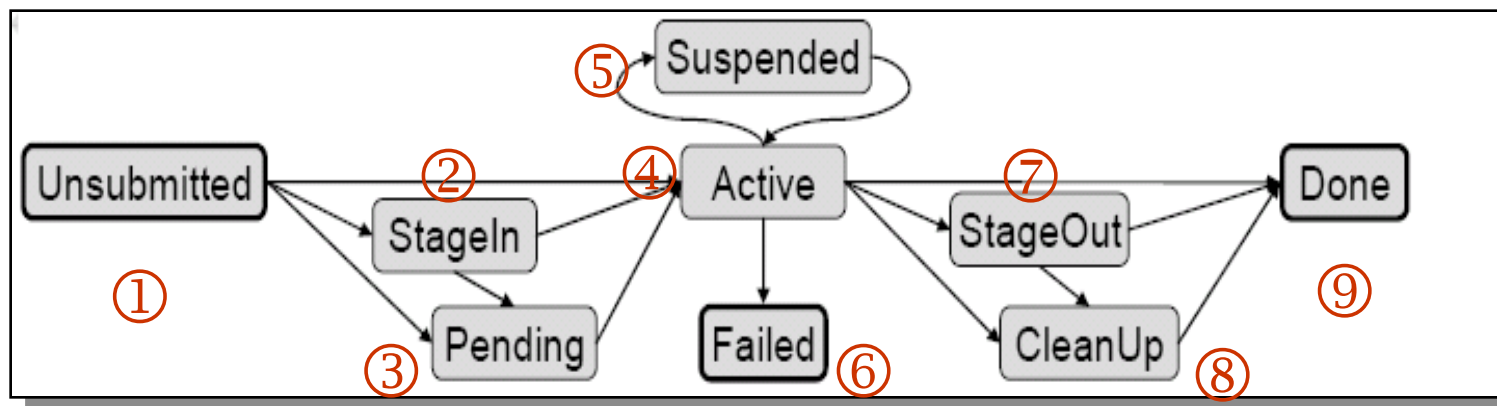
- GT4 fornisce una serie di web service per eseguire, monitorare e terminare l'esecuzione di job su una Griglia.
- Essi compongono il WS_GRAM che estende le funzionalità del componente GRAM (Grid Resource Allocation and Management) prima dell'uso dei web services.
- Interfacce a linee di comandi e API in Java, C e Python.

Esecuzione Remota di Programmi

- Il **WS_GRAM** permette la gestione dell'esecuzione remota di programmi in maniera affidabile, con meccanismi di monitoring, gestione delle credenziali, staging dei file da usare, e interazione con gli scheduler locali.
- Gestisce:
 - Gli eseguibili
 - Lo stato del processo
 - L'I/O
 - Il controllo remoto
 - Gli scheduler locali
 - Il monitoraggio dei processi

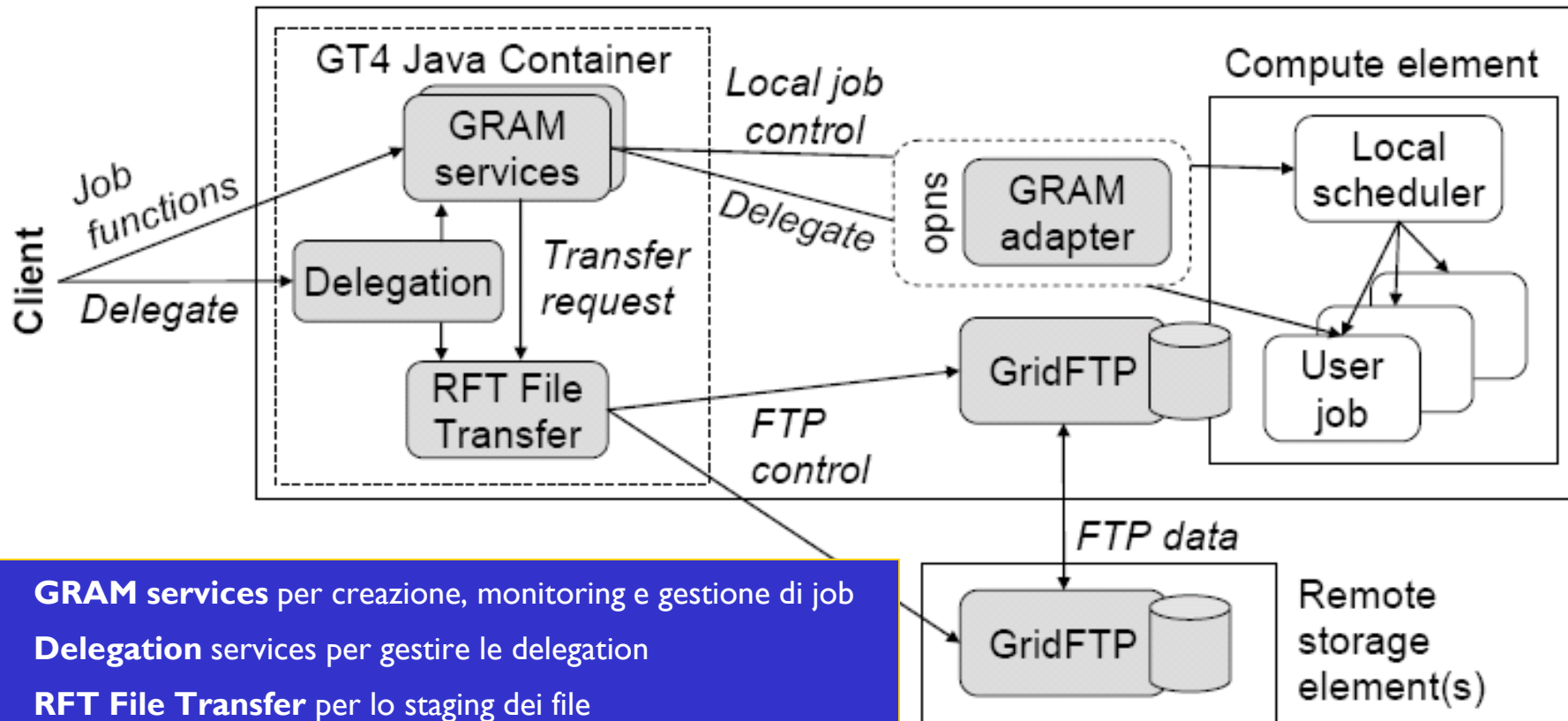
Stato dei Job

- Il GRAM gestisce i job e le loro transizioni di stato



Stato	Descrizione
Unsubmitted	Di prossima esecuzione
StageIn	In attesa che i file vengano resi disponibili per l'esecuzione
Pending	In attesa che lo scheduler locale ne decida l'esecuzione
Active	In esecuzione
Suspended	Job sospeso
StageOut	Esecuzione completata e file di output resi disponibili
CleanUp	Rimozione delle risorse usate
Done	Job completato con successo
Failed	Job fallito

Implementazione del GRAM



GRAM services per creazione, monitoring e gestione di job

Delegation services per gestire le delegation

RFT File Transfer per lo staging dei file

GRAM adapter per l'interazione con lo scheduler locale

GridFTP per eseguire lo staging dei dati

Politiche di Sicurezza di Globus

Le politiche di sicurezza di Globus sono basate sui seguenti principi.

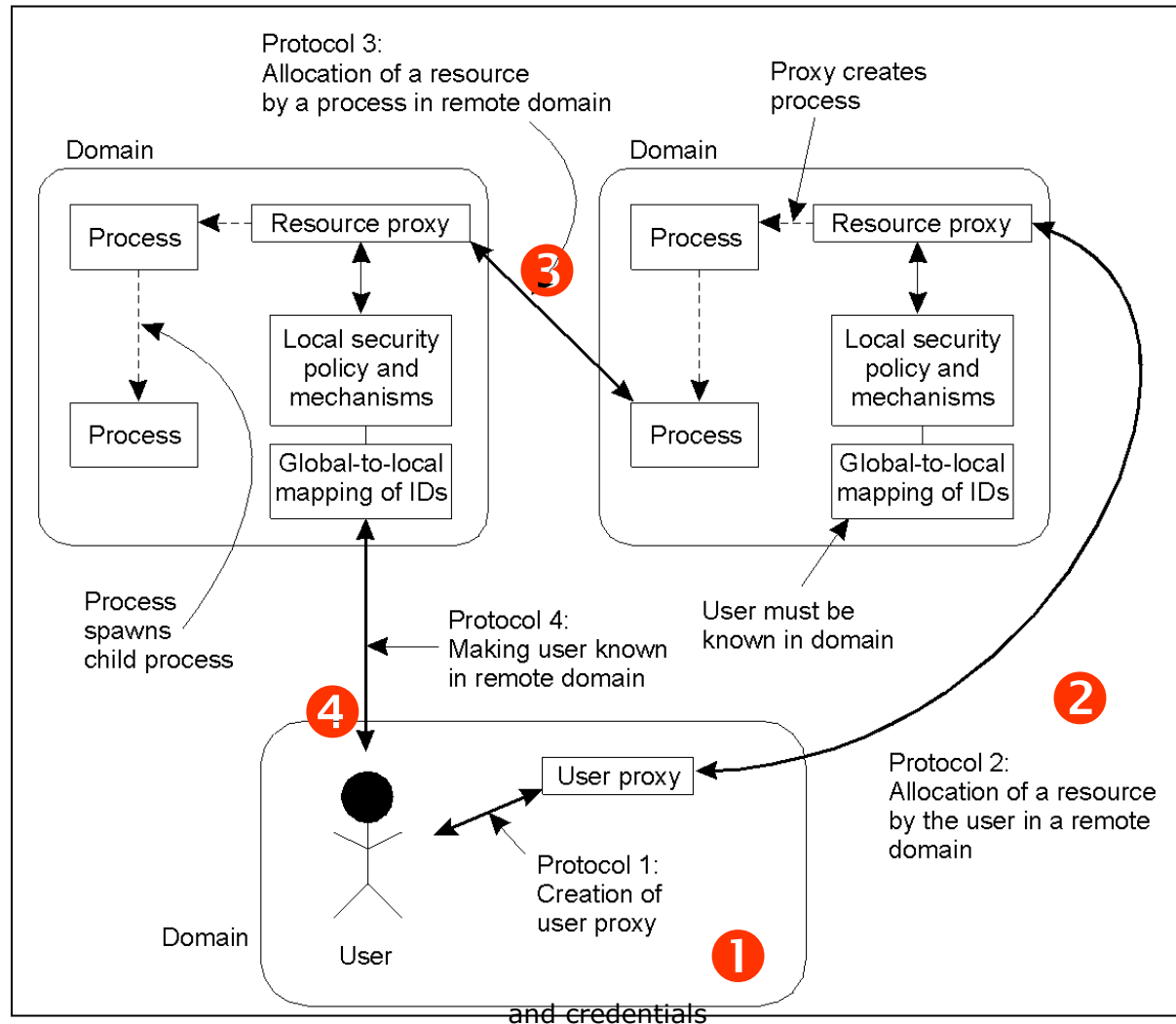
- 1. Il sistema consiste di diversi domini amministrativi.**
- 2. Le operazioni locali sono soggette solo a politiche di sicurezza locali.**
- 3. Le operazioni globali richiedono che il richiedente sia riconosciuto in tutti i domini interessati alle operazioni.**
- 4. Le operazioni tra entità in domini differenti richiedono la mutua autenticazione.**
- 5. L'autenticazione globale sostituisce la autenticazione locale.**
- 6. Il controllo degli accessi alle risorse è soggetto alle politiche locali.**
- 7. Gli utenti possono delegare loro diritti ai processi.**
- 8. Un gruppo di processi nello stesso dominio può condividere le credenziali.**

Globus Security Architecture (1)

- La *Globus security architecture* consiste di entità come utenti, processi, user proxies, e resource proxies.
- Uno **user proxy** è un processo che ha il permesso di agire per conto di un utente per un limitato periodo di tempo.
- Un **resource proxy** è un processo che in un dominio è usato per tradurre operazioni globali su una risorsa in operazioni locali che rispettano la politica di sicurezza del dominio.

Globus Security Architecture (2)

Diagramma della
*Globus security
architecture*
e dei protocolli.



Griglie Computazionali

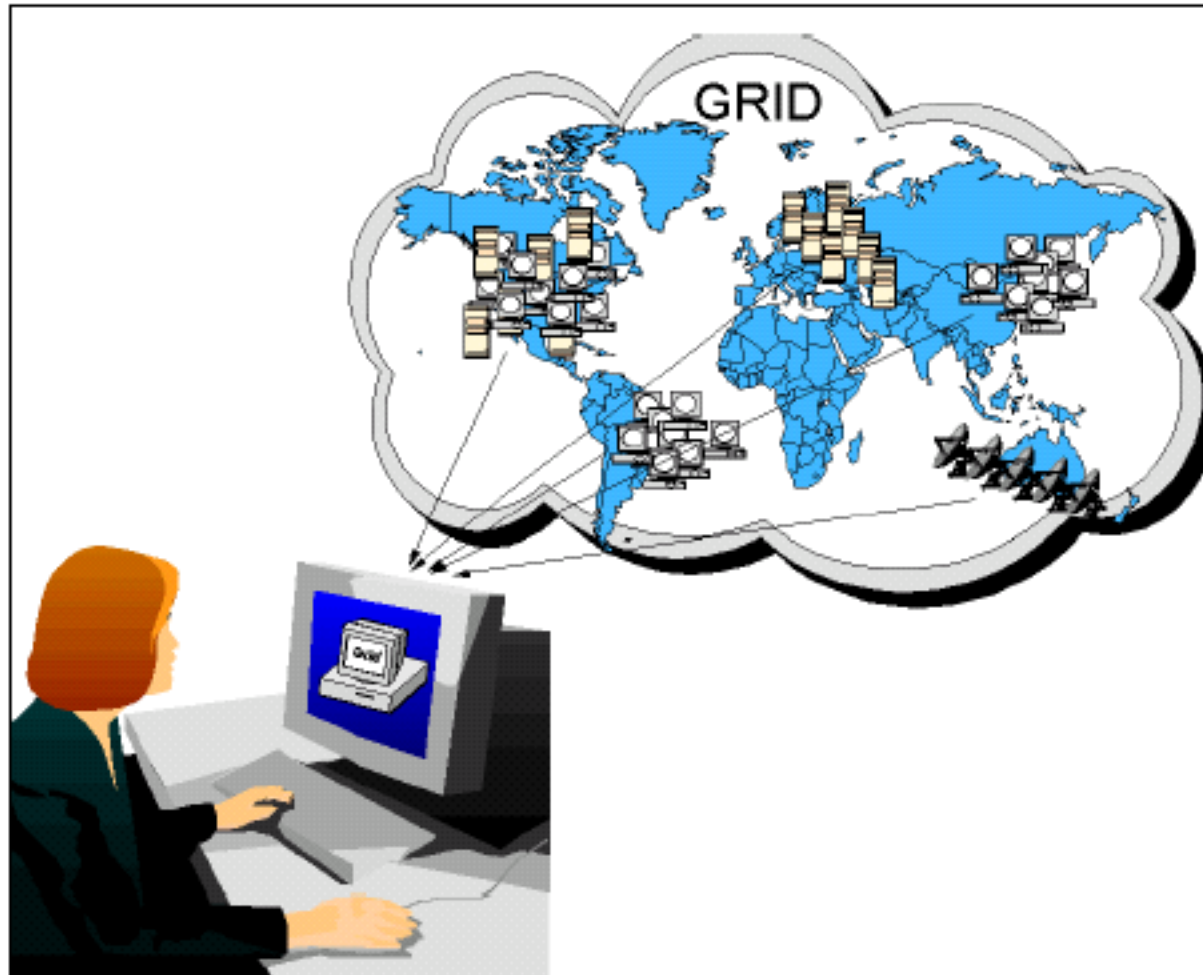


Figure 1-1 The grid virtualizes heterogeneous geographically disperse resources

Griglie Computazionali

I sistemi di Grid computing possono essere realizzati a più livelli:

- **Intragrids**
- **Extragrids**
- **Interagrids**

Ognuno di questi modelli si estende in domini e organizzazioni diverse. Tutti potrebbero essere integrati in una **World Wide Grid**.

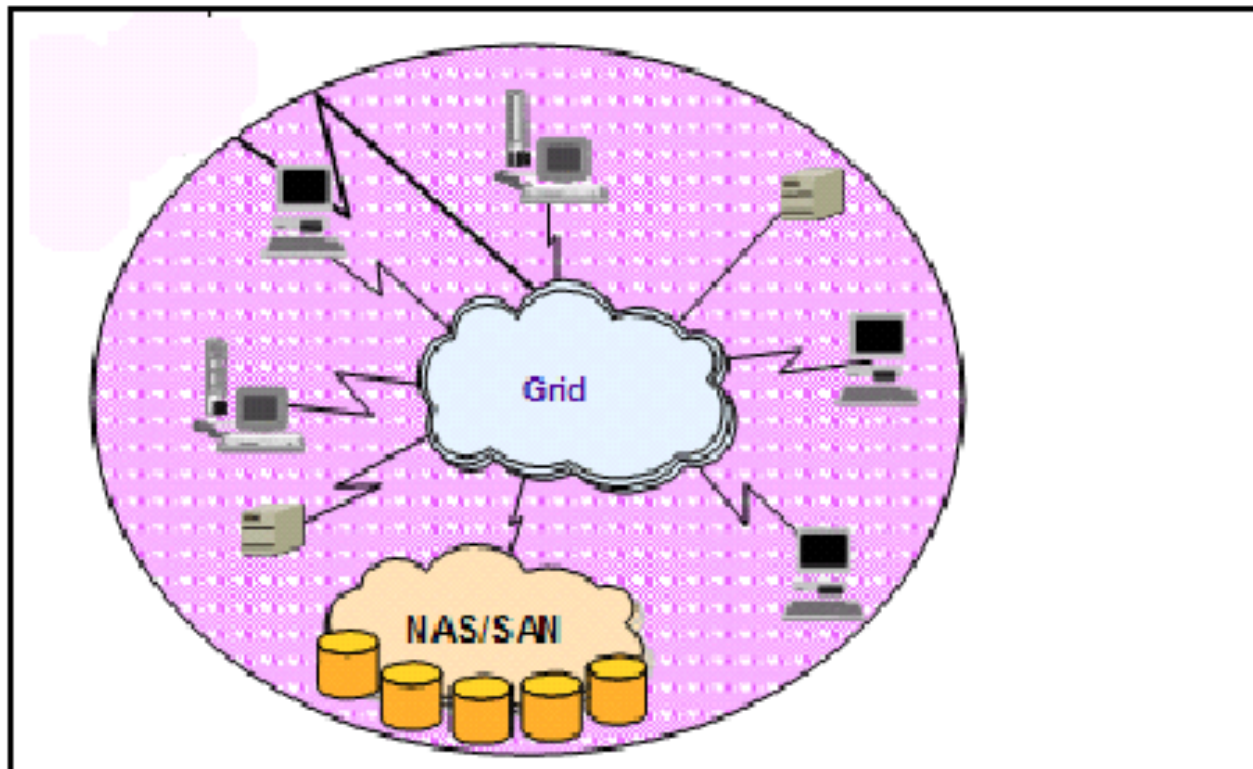
Griglie Computazionali

I sistemi di Grid computing possono essere realizzati a più livelli:

Intragrids

- Singole organizzazioni
- Nessuna integrazione con partner
- Es: Un singolo cluster

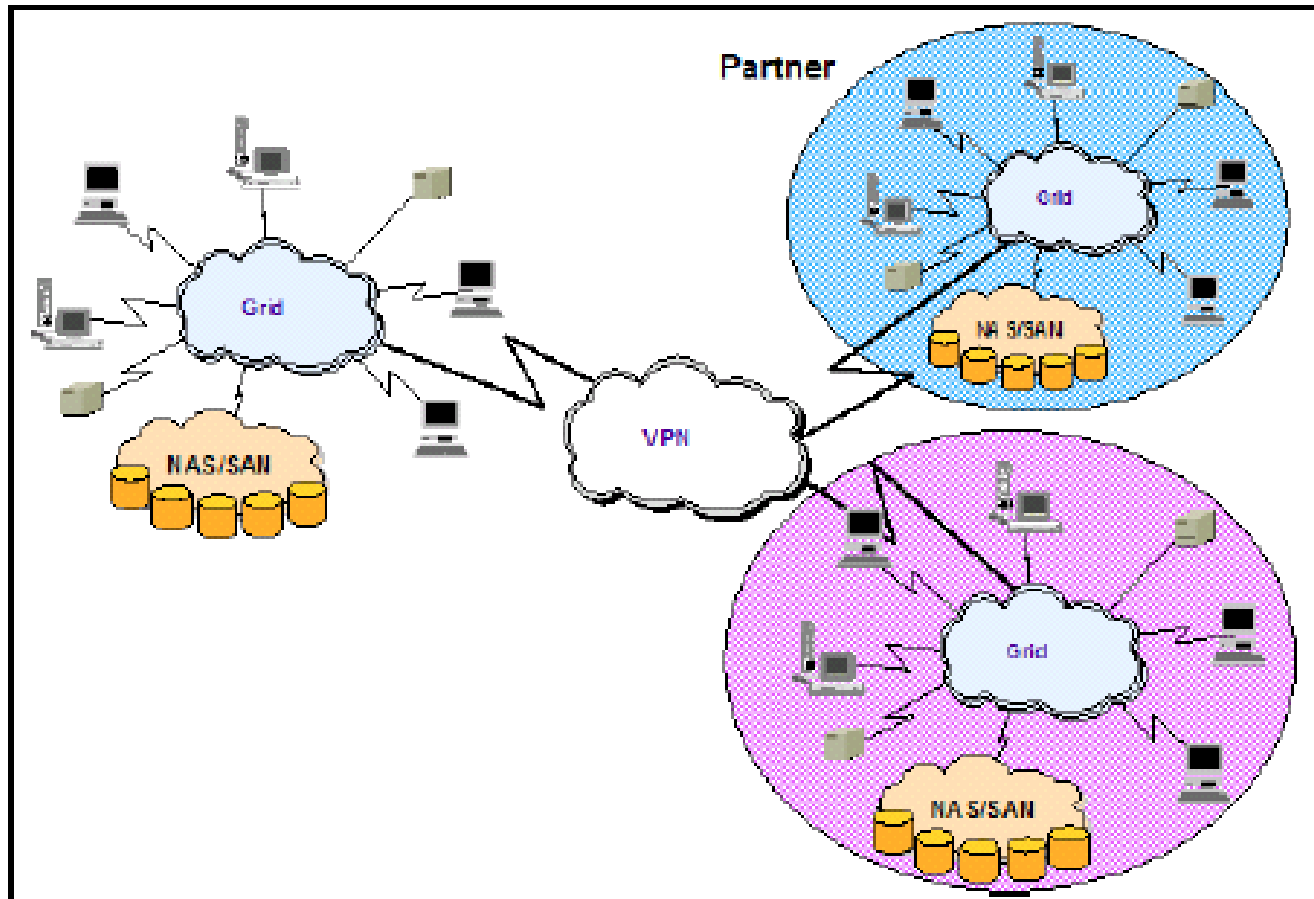
Griglie Computazionali : Intragrids



Extragrids

- Più organizzazioni
- Integrazioni di Partner
- Più computer (clusters, PC, ecc.)

Griglie Computazionali : Extragrids

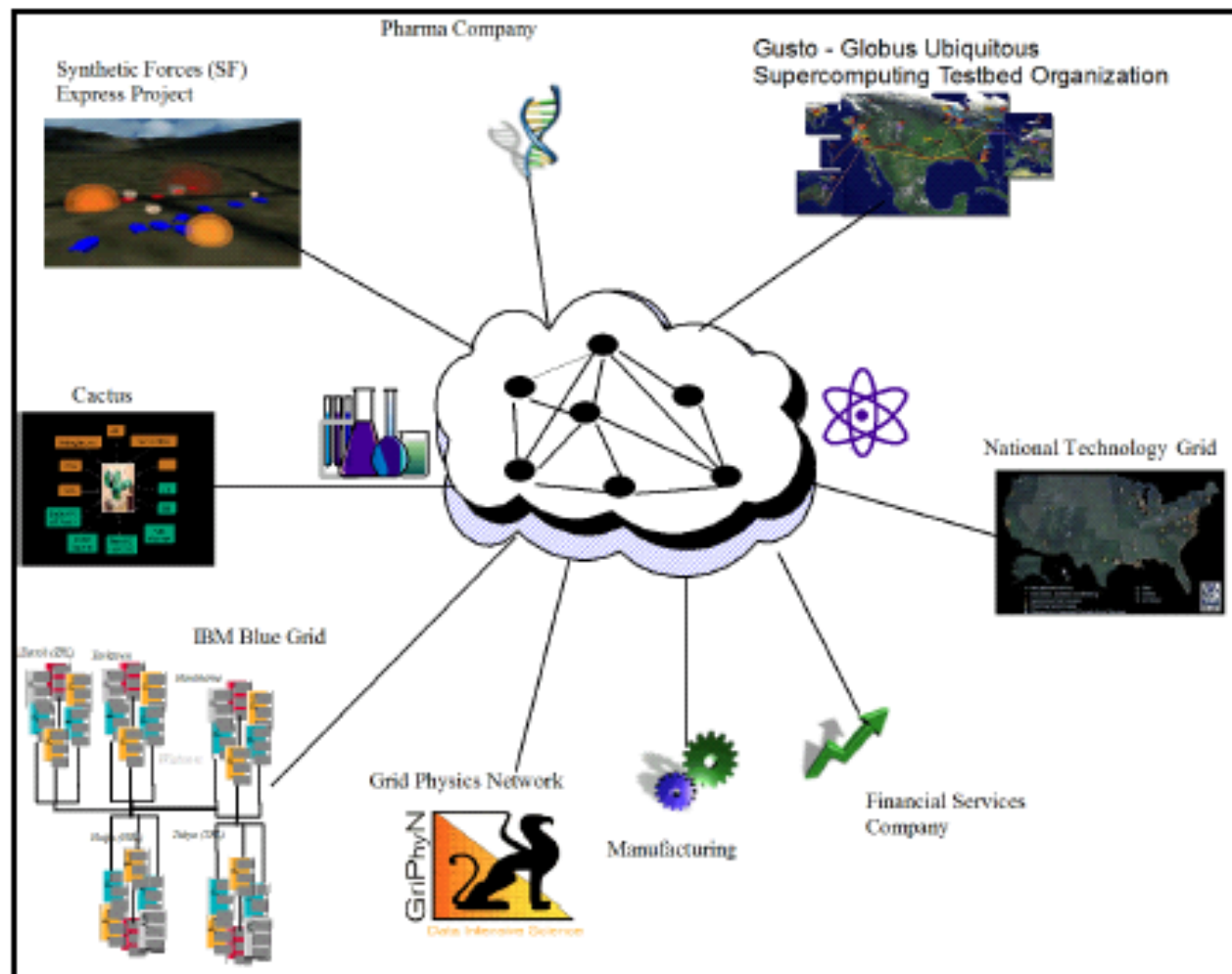


Griglie Computazionali : Intergrids

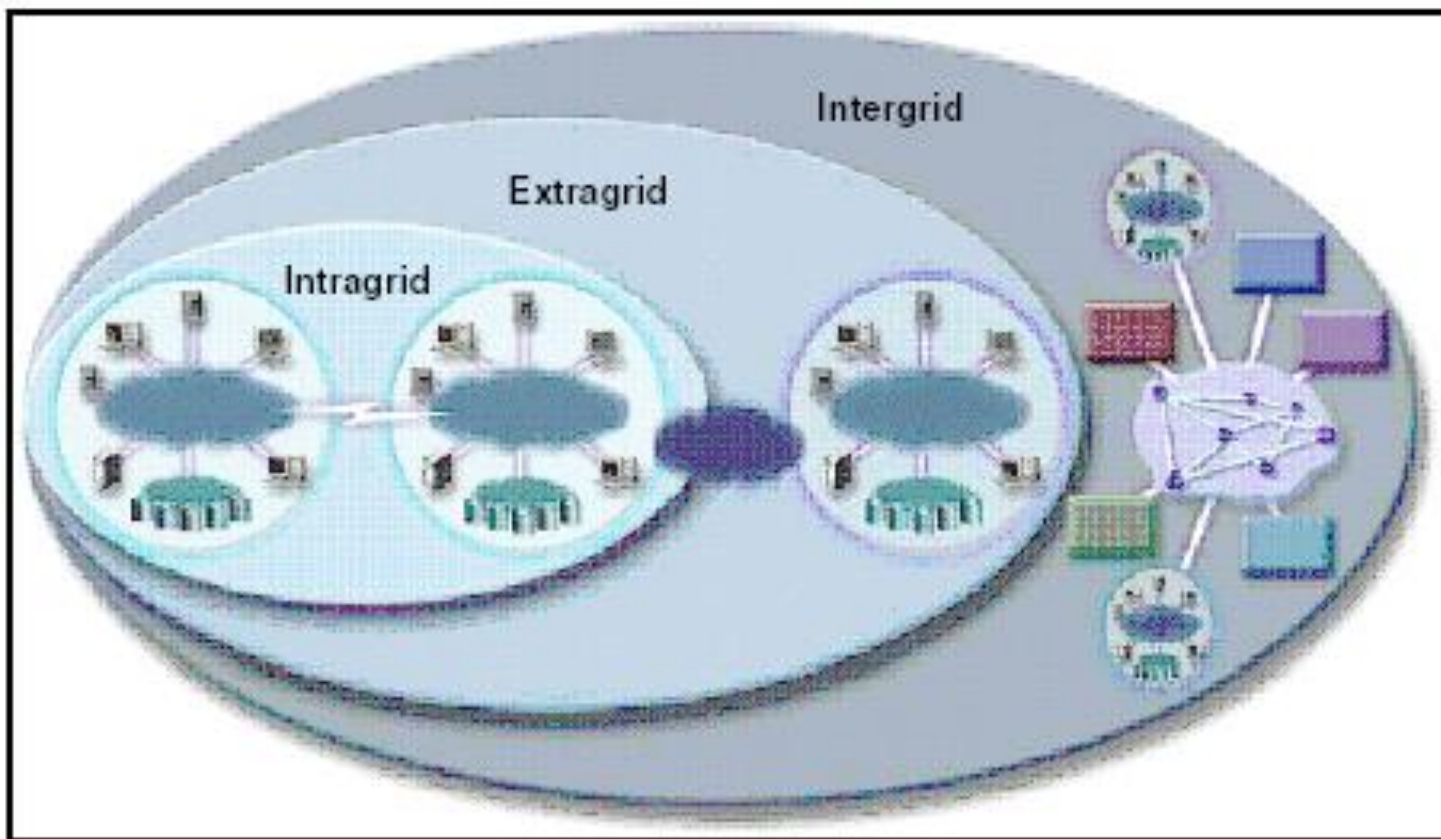
Intergrids

- Molte organizzazioni
- Multi Partners
- Multi computer (PC, clusters, Intragrids)

Griglie Computazionali : Intergrids



Griglie Computazionali : Livelli di Integrazione



Verso i Sistemi di Ubiquitous Computing

Fino a 30 anni fa:

Un computer per molti utenti.



Fino a pochi anni fa:

Un computer per ogni utente.



Oggi e in futuro:

Tanti computer per ogni utente.



Verso i Sistemi di Ubiquitous Computing

Oggi:

Internet connette tutti i computer sulla terra e nel sistema solare.

Futuro:

Ogni sistema di elaborazione, ogni dispositivo di comunicazione e ogni oggetto o manufatto (smart objects) potrà essere connesso indipendentemente da dove esso si trovi.

I sistemi Cloud (insieme a Edge e Fog) costituiscono il backbone di questo sistema.