Il Grid computing: oltre Internet

(L. Merola)

L'infrastruttura Grid

Il Grid computing e quindi l'infrastruttura che esso utilizza fornisce una soluzione al problema del calcolo distribuito intensivo e ad alte prestazioni tramite l'accesso continuo ed organizzato ad ingenti risorse di vario tipo (calcolatori, banche dati, librerie software) sparse su scala geografica ma interconnesse fra loro tramite reti di comunicazione dati a larga banda (ad alta velocità) da parte di grandi comunità internazionali composte da migliaia di persone disperse geograficamente ma cooperanti per uno stesso obiettivo (scientifico, industriale, di ricerca, commerciale, ecc.).

Le risorse sono condivise in modo trasparente all'utente finale, in analogia – in un certo senso – alle reti di distribuzione di altri servizi quali l'energia elettrica o il telefono. In effetti il termine Grid trae origine proprio dall'inglese *electric power grid* che significa appunto rete elettrica.

Ad esempio, così come l'utente inserisce la spina di un'apparecchiatura (da una semplice lampadina ad un elettrodomestico o ad un qualunque dispositivo elettrico) nella presa elettrica senza sapere né dove siano le centrali né come sia trasportata l'energia elettrica ma in compenso pretende che il servizio sia affidabile, continuo e rispondente – entro certi limiti – alle sue richieste, così l'utente della Grid inserisce il suo sistema di interfacciamento alla rete (tipicamente un PC), senza sapere dove siano dislocate le risorse di calcolo e di immagazzinamento dei dati ma pretende anche in questo caso una qualità di servizio che non causi interruzioni ed in generale risponda alle sue richieste di utilizzo anche intensivo per periodi limitati.

In tale modo si realizza la virtualizzazione delle risorse e degli utenti, che si costituiscono in organizzazioni virtuali omogenee per attività o interessi. Naturalmente come per l'energia elettrica (o il telefono o altri servizi collettivi) occorrono le attrezzature e le organizzazioni che gestiscano e controllino il buon funzionamento del servizio, anche nel caso dell'infrastruttura Grid si vanno sempre di più stabilendo le regole di funzionamento e consolidando le organizzazioni nazionali ed internazionali che se ne occupino.

Il grande vantaggio del Grid computing consiste nella disponibilità e nell'ottimizzazione dell'uso di grandi risorse di calcolo e dati che nessuna singola organizzazione sarebbe altrimenti in grado di acquistare e gestire da sola e che ha perciò la possibilità di utilizzare solo quando ne ha realmente bisogno con la ridondanza necessaria e la qualità ed affidabilità del servizio, continuamente monitorato grazie ad appositi centri di gestione e controllo.

L'architettura Grid si basa su piattaforme abilitanti hardware *multivendor* e su un'infrastruttura di software intermedio, chiamato appunto *middleware*, che interfaccia l'hardware con le applicazioni e contiene in sé servizi di gestione delle risorse sia locali (servizi *core*) sia mondiali (servizi *collective*) basati sul concetto di risorse virtuali condivisibili dalle comunità omogenee denominate organizzazioni virtuali prima citate.

Obiettivi dei progetti Grid

I progetti Grid si sono sviluppati traendo grande beneficio dall'esperienza della nascita e dello sviluppo del web e dall'esplosione del fenomeno Internet. Si ricorderà che il web nacque alla fine degli anni '80 al CERN di Ginevra – il centro europeo per la fisica delle particelle - per consentire a centinaia o migliaia di ricercatori di tutto il mondo di comunicare fra loro e di scambiarsi documenti e dati tramite una semplice interfaccia software (*browser*) appositamente realizzata.

Oggi si può dire che stiamo assistendo al passaggio dal World Wide Web, il sistema informativo globale ormai diffusissimo in tutto il mondo che attualmente tratta dati multimediali (testi, immagini, filmati) e le cui capacità di interattività sono ancora tutte da esplorare, al World Wide Grid, il sistema di calcolo distribuito e di gestione di grandi moli dati su scala mondiale,

basato ancora su semplici interfacce utente che consentono l'accesso ai servizi Grid con una qualità di servizio appositamente garantita. Il nuovo "Internet" sarà basato sui "servizi" che potranno essere offerti a comunità diversificate, scientifiche e non, e per questo motivo si parla sempre di più di "Internet dei servizi".

Risale al 1999 il testo fondamentale di Ian Foster (Univ. Chicago) e Carl Kesselman (Univ. California), "The Grid Blueprint for a New Computing Infrastructure". Non per caso dunque i primi grandi progetti Grid si svilupparono negli USA in quel periodo, dopo i numerosi tentativi pre-Grid (ad es. il sistema Condor degli anni '80), e si diffusero rapidamente in Italia ed in Europa grazie a finanziamenti nazionali specifici di Enti di ricerca (tra quelli nazionali vanno citati fra gli altri l'INFN, l'INAF, il CNR) e ministeriali (in Italia i FIRB, i PON, fra gli altri), e i progetti nazionali quali ad esempio INFN-GRID, Grid.It, SPACI. Tra i progetti europei (molti dei quali nell'ambito dei programmi quadro dell'UE) ricordiamo ad esempio DataGrid (EDG), DataTag, DEISA Distributed European Infrastructure for Supercomputing Appications), AstroGrid, LCG (LHC Computing Grid), EGEE (Enabling Grid and E-sciencE). E' notevole che l'UE abbia anche avviato progetti per la diffusione della Grid in Cina (EUChinaGrid), in India (EUIndiaGrid), nei Paesi del mediterraneo (EUMedGrid), in America Latina (EELA).

Lo scopo principale è la realizzazione di una IGI (*Italian Grid Infrastrucure*) che si inserisca nella EGI (*European Grid Infrastructure*) che si colleghino fra loro e siano interoperabili con l'infrastruttura Grid americana e quella asiatica. In effetti si parla, oggi, di e-infrastrutture a livello mondiale che combinano Grid e la rete Internet a larga banda che è basata sulle più moderne tecnologie di trasmissione dati; si pensi che siamo passati da poche migliaia di bit al secondo (inizi anni '90) alle decine di miliardi di bit al secondo di oggi.

In Italia riveste un ruolo particolarmente importante il Consortium GARR (acronimo dell'originario Gruppo Armonizzazione delle Reti per la Ricerca) che da quasi trent'anni si occupa e si preoccupa di concepire e gestire l'ossatura (il *backbone*) delle linee di comunicazione dati per l'Università e gli Enti di ricerca, in collegamento con i gestori delle telecomunicazioni che operano sul territorio, lasciando alle singole realtà locali la progettualità e la realizzazione delle reti locali e interne.

La diffusione dell'esperienza Grid ha raggiunto ormai livelli eccellenti e la citazione di tutte le realtà diventerebbe troppo lunga e necessariamente non esaustiva. Interessante è notare che in molti progetti sono protagonisti i grandi Consorzi o Centri di Supercalcolo (ad es. DEISA) ed è anche consolidata la partecipazione attiva di aziende e imprese e sicuramente il loro coinvolgimento sarà sempre maggiore, quanto più si diffonderà la consapevolezza delle potenzialità della Grid.

Nel recente passato il MiUR ha cofinanziato quattro progetti cofinanziati tramite l' Avviso 1575/2004 – Sistemi di calcolo ad alte prestazioni - si inseriscono nell'ambito del PON Ricerca 2000-2006 per le regioni dell'Obiettivo 1 (Basilicata, Calabria, Campania, Puglia, Sardegna e Sicilia) – cioè le regioni in ritardo di sviluppo - ed in particolare nell'ambito dell'Asse strategico II "Rafforzamento e apertura del sistema scientifico a di alta formazione meridionale". Due di essi sono progetti che vedono soggetti attuatori campani: il progetto **S.Co.P.E.** (**S**istema **Co**operativo distribuito ad alte **P**restazioni per **E**laborazioni Scientifiche Multidisciplinari), dell'Università di Napoli Federico II, a cui collabora in modo particolarmente significativo la Sezione INFN di Napoli, e il progetto **CRESCO** (**C**entro Computazionale di **R**ic**E**rca sui **S**istemi **CO**mplessi), presentato dall'ENEA, che ha come obiettivo la realizzazione, presso il Centro Ricerche ENEA di Portici (NA), di un importante Polo di calcolo multidisciplinare.

Gli altri due progetti sono il progetto **CYBERSAR** (**CYBER**infrastructure per la ricerca scientifica e tecnologica in **SAR**degna) gestito dal consorzio COSMOLAB che include il CRS4 (Centro per gli Studi Avanzati, la Ricerca e lo Sviluppo in Sardegna), l'INAF (Istituto Nazionale di Astrofisica), l'INFN (Istituto Nazionale di Fisica Nucleare) l'Universita' di Cagliari, l'Universita' di Sassari, la società Nice s.r.l. e la società Tiscali S.p.a. ed il progetto **PI2S2** (**P**rogetto per l'**I**mplementazione e lo **S**viluppo di una e-**I**nfrastruttura in **S**icilia basata sul paradigma delle

GRID) gestito dal Consorzio COMETA che include INFN, INAF, INGV, UniCT, UniME, UniPA ed il Consorzio SCIRE.

I quattro progetti si inseriscono perfettamente nel quadro nazionale ed europeo dei progetti di Grid computing. Come valore aggiunto essi stanno lavorando proprio in questo periodo a rendere interoperabili le proprie infrastrutture ed applicazioni. Ciò si realizzerà attraverso l'adozione di uno standard comune di comunicazione e di servizi di base comuni e dovrà prevedere anche l'interoperabilità e la facile integrazione con le altre infrastrutture di calcolo distribuito nazionali ed internazionali basate sulla tecnologia Grid, nella consapevolezza che il successo di un'infrastruttura risiede nella sua capacità di interoperare con molte Grid che adottano standard diversi.

I quattro progetti hanno anche l'ambizione di promuovere sinergie fra comunità scientifiche ed accademiche che si occupano degli stessi o di analoghi temi di ricerca e che troveranno nell'infrastruttura Grid uno strumento di cooperazione di grandi potenzialità; intendono altresì porsi come componente fondamentale dell'asse portante dell'infrastruttura meridionale al fine di promuovere ed elevare anche, in prospettiva, la competitività delle Imprese nel Mezzogiorno.

Potenzialità di sviluppo del Grid computing.

Le Grid sono nate in ambito scientifico ma le loro potenzialità, esattamente come è avvenuto per il web, stanno interessando sempre di più le industrie, le imprese, le amministrazioni, gli enti locali e nazionali, le organizzazioni governative, commerciali, finanziarie.

Molti sono i settori applicativi nei quali vengono immagazzinate grandi moli dati (o reali o ottenuti attraverso modelli di simulazione) e per i quali è necessaria la loro elaborazione, talvolta in tempo reale. Basti pensare alle simulazioni ed alla modellistica per le scienze fisiche, chimiche, matematiche, la medicina, la biologia, l'ingegneria, le scienze umane e sociali, basti pensare all'analisi dei dati provenienti da satelliti o da sensori terrestri per la meteorologia, la tutela e la conservazione del territorio e dell'ambiente, dei beni culturali, della salute (ad es. l'analisi a distanza dei dati clinici e medici), la protezione civile (ad es. la previsione o almeno il pronto intervento in seguito a disastri quali terremoti, inondazioni, incendi), la sicurezza in generale (ad es. contro attacchi terroristici di qualunque genere).

E' chiaro che la disponibilità di risorse di calcolo quasi illimitate, grazie al paradigma del Grid computing, in grado di gestire grandi moli di dati digitali in tempi rapidi e in totale sicurezza per le azioni conseguenti, costituisce un vero valore aggiunto non reperibile nei tradizionali sistemi distribuiti che soffrono di rigidità, limitatezza e difficoltà di utilizzo, e non offrono ridondanza di risorse e qualità garantita del servizio.

Va infine sottolineato che Grid non è solo un"'infrastruttura"; è infatti anche una tecnologia nuova che coinvolge tutti e tre gli strati fondamentali: *hardware*, *midlleware* e *software.* In particolare comporterà lo sviluppo di interfacce utente ad alto livello (ad esempio portali web innovativi) ma imporrà anche un modo diverso di progettare e pensare nuovi algoritmi, nuove librerie software, nuove tecniche di trattamento e manipolazione dei dati, di simulazione e di analisi dati che rendano realmente integrabili ed interoperabili le applicazioni in una piattaforma globale di supercalcolo distribuito.

In questo sforzo i futuri contributi delle Regioni attraverso i POR, del MUR e del MSE attraverso i PON, dell'Europa attraverso i programmi quadro, dovranno marciare tutti nella stessa direzione per evitare sprechi e mettere a fattor comune esperienze e competenze. Importante sarà in particolare il nuovo Programma Operativo Nazionale "Ricerca e Competitività" 2007-2013 che si inserisce nel Quadro Strategico Nazionale che l'Italia ha presentato all'UE e che prevede una programmazione specifica che si realizzerà nelle regioni "Convergenza" del Sud e prevederà Azioni di Sistema tese anche all'integrazione delle azioni dei PON con quelle dei POR e a stabilire forti legami fra le esperienze del Nord e quelle del Sud.