

Comput-ER: l'infrastruttura e i servizi di calcolo e archiviazione distribuiti per le comunità scientifiche e non in Emilia Romagna

Marco Bencivenni (INFN-CNAF), Daniele Cesini (INFN-CNAF), Paolo Veronesi (INFN-CNAF)
in rappresentanza della collaborazione DUCK

Le tecnologie del calcolo distribuito, quali la Grid e più recentemente anche il Cloud Computing, hanno dimostrato vantaggi non solo in termini economici, ma soprattutto dal punto di vista della promozione di collaborazioni e dello scambio scientifico fra le varie comunità scientifiche. La grande quantità di dati prodotti dalla ricerca nei campi della Fisica delle Alte Energie, astronomia/astrofisica, biologia, medicina molecolare, fisica della terra, climatologia, oceanografia, ecc., pone in modo sempre più urgente la necessità di razionalizzare la memorizzazione di questi dati e la loro condivisione da parte delle comunità scientifiche e tra queste ultime e le realtà istituzionali presenti sul territorio. Questi paradigmi di calcolo e di accesso ai dati hanno dato un grande impulso alla ricerca producendo importanti risultati in brevissimo tempo. Inoltre l'emergere di aree di nuovi interessi, quali la meta-genomica, sta suggerendo che l'integrazione fra discipline differenti possono aiutare nell'affrontare problemi ancora senza risposta. E' altresì evidente che la condivisione di middleware e infrastrutture comuni stimola iniziative scientifiche multi-disciplinari, permette il trasferimento di tecnologie più all'avanguardia da un campo all'altro consentendo la condivisione delle innovazioni, l'abbassamento dei costi dello sviluppo e favorisce inoltre l'eliminazione di duplicazioni.

Le iniziative sul calcolo distribuito europee e internazionali sono state e sono tuttora sostenute dal VI programma quadro (DataGRID, DataTAG) e poi VII programma quadro europeo (EGEE, EGIDS, PRACE), da vari progetti nazionali (GRid-IT, FIRB, PON) e regionali; inoltre il Ministero dell'Università e della Ricerca ha inserito Grid fra le piattaforme abilitanti e sostiene il consolidamento della Grid Italiana, IGI (Italian Grid Infrastructure). Queste iniziative hanno coinvolto gruppi di varie comunità scientifiche e molti di questi sono operanti in Emilia Romagna. È emersa quindi l'esigenza di estendere le nuove tecnologie per il calcolo distribuito in modo più capillare nel territorio della Regione, consolidando l'attuale coordinamento in atto, coinvolgendo altre entità nel panorama dell'università e della ricerca, al fine di sviluppare nuovi scenari applicativi da estendere a diversi settori quali per es. la sanità, la pubblica amministrazione etc.

Rete ad alta velocità e certificati x509 sono gli elementi essenziali per l'accesso all'infrastruttura Grid italiana, gestita dall'Italian Grid Infrastructure (IGI), e completamente interconnessa con quella Europea EGI. Gli enti, Università e istituti cui queste comunità afferiscono hanno l'indubbio vantaggio di essere collegate mediante la rete ad alta velocità fornita dal consorzio GARR e possono inoltre usufruire di un set di servizi base grazie all'appartenenza alla federazione IDEM. Gli enti che fanno parte di questa federazione hanno quindi a disposizione sistemi di Autenticazione e autorizzazione comuni con la possibilità di richiedere certificati di tipo x509 rilasciati da Terena e validi per l'utilizzo nell'infrastruttura di Grid italiana ed europea.

E' quindi nata una collaborazione, denominata DUCK (Distributed Unified Computing for Knowledge), per lo svolgimento di attività di ricerca e sviluppo nel campo dei servizi e delle

infrastrutture di Grid/Cloud, finalizzate al potenziamento e alla maggiore funzionalità delle infrastrutture e dei servizi esistenti nell'ambito della Regione Emilia Romagna, nonché alla partecipazione, attraverso l'integrazione delle proprie risorse computazionali e di dati, al consolidamento ed all'espansione dell'infrastruttura al servizio di altre componenti della regione e aziende private.

Le organizzazioni che partecipano alla collaborazione sono:

- Istituto Nazionale di Fisica Nucleare (CNAF e sezione di Bologna);
- Istituto Nazionale di Geofisica e Vulcanologia (sezione di Bologna);
- Centro EuroMediterraneo per i Cambiamenti Climatici;
- Centro Interdipartimentale di Ricerca sul Cancro "GIORGIO PRODI" (Università di Bologna);
- Dipartimento di Biologia Evoluzionistica Sperimentale dell'Università di Bologna;
- CeIRG Centro Interdipartimentale di Ricerche Genomiche dell'Università di Modena e Reggio Emilia;
- Area di Ricerca del CNR di Bologna (CNR/INAF);
- Dipartimento di Fisica dell'Università di Ferrara;
- Dipartimento di Fisica dell'Università di Bologna;
- Dipartimento di Fisica dell'Università di Parma;
- Dipartimento di Scienze dell'Informazione dell'Università di Bologna.

La partecipazione alla collaborazione avviene utilizzando le strutture di ricerca fornite dalle organizzazioni partecipanti e inserite nell'infrastruttura Grid italiana. Le proposte di progetti o piani di lavoro collaborativo contengono le specifiche attività da svolgere, gli obiettivi da realizzare, i termini e le condizioni di svolgimento, le risorse umane e strumentali da impiegare e messe a disposizione dalle parti, la definizione degli oneri finanziari relativi alle singole attività tecnico-scientifiche e delle loro modalità di erogazione, il regime dei risultati scientifici e la disciplina dei diritti di proprietà intellettuale applicabile.

Nell'ambito di questa collaborazione, sono svolte attività finalizzate a:

- consolidare, rafforzare ed espandere l'infrastruttura di calcolo regionale partecipando ai bandi della Regione, a bandi MIUR e a quelli europei con progetti comuni;
- espandere gli scenari applicativi ad altri settori della ricerca, della sanità e della P.A. per sviluppare nuove applicazioni che traggano vantaggio dall'uso delle tecnologie del calcolo distribuito;
- condividere, diffondere e pubblicizzare le conoscenze e le tecnologie nel campo del calcolo distribuito quali Grid e Cloud;
- promuovere il trasferimento tecnologico;
- ottimizzare attraverso un efficace coordinamento la gestione della infrastruttura ed evolverla secondo gli standard nazionali ed europei;
- promuovere l'organizzazione di convegni, seminari, scuole, gruppi di lavoro interdisciplinari per l'uso dell'infrastruttura Grid/Cloud.

A supporto delle attività è stata predisposto un sito Web (<http://www.comput-er.it/>) e un'infrastruttura composta dalle strutture di ricerca delle organizzazioni partecipanti, riportata in Figura 1 (in viola i centri di calcolo in via di certificazione).

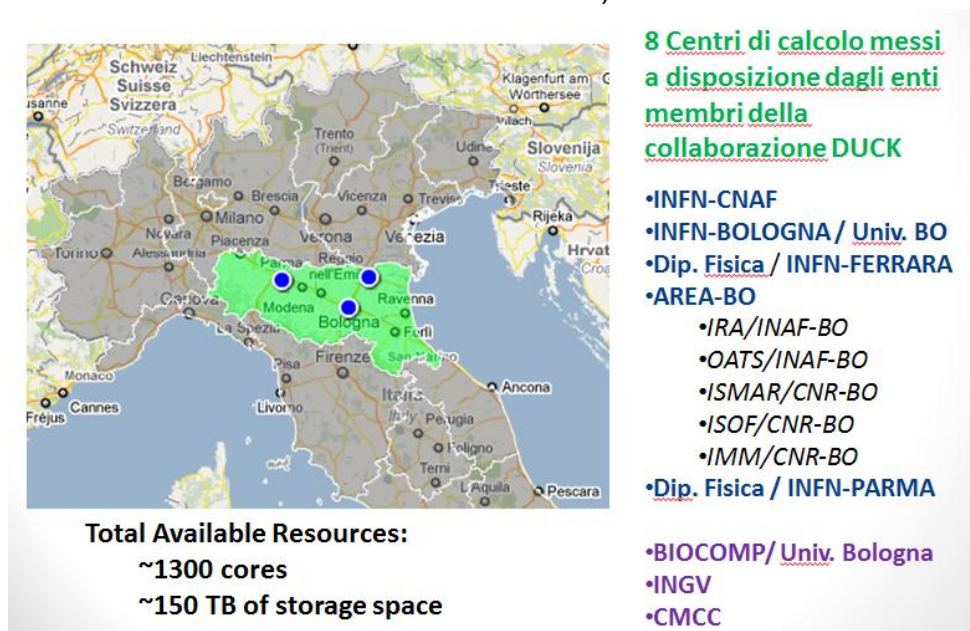


Figura 1. L'infrastruttura Grid a supporto delle attività della collaborazione

Le risorse di calcolo e storage a disposizione sono circa 1300 core e 150TB di spazio disco, utilizzabili in maniera opportunistica per le diverse attività della collaborazione.

L'infrastruttura prevede inoltre una serie di servizi Grid di alto livello opportunamente configurati e completamente integrati nell'infrastruttura di calcolo Italiana ed Europea. In particolare, è stata creata una Virtual Organization (comput-er.it) nella quale vengono inserite le persone che partecipano alle diverse attività e tale VO è stata abilitata nei diversi centri di calcolo. Inoltre, sono stati configurati un catalogo dei dati (LFC) e un pool do WMS per la sottomissione dei job.

L'interfacciamento, il supporto agli utenti e agli amministratori dei siti Grid locali, i sistemi di monitoring e accounting utilizzati dalla collaborazione e i servizi di alto Grid di alto livello sono gestiti dal gruppo Grid Operation, presso il CNAF, nell'ambito delle attività di supporto e controllo dell'infrastruttura Grid italiana. Il gruppo inoltre collabora con i ricercatori degli altri istituti per analizzare le applicazioni proposte in modo da individuare gli scenari più opportuni per eseguire le applicazioni in ambiente Grid.

I risultati di questa collaborazione, nata nel Maggio 2010, sono stati presentati in un workshop tenutosi nel Novembre 2010 (<https://agenda.italiangrid.org/conferenceDisplay.py?confId=364>) e con il quale la collaborazione si è presentata alle istituzioni regionali.

Sono inoltre state prodotte una serie di presentazioni e articoli, sia a livello nazionale che internazionale, tra cui si segnala la partecipazione all'EGI User Forum 2010

(<http://uf2011.egi.eu/>) nei quali sono stati presentate le applicazioni portate in Grid e i benefici ottenuti utilizzando questo tipo di infrastruttura.