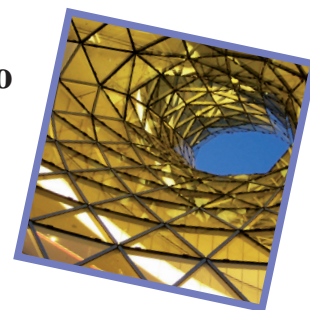


Comput-ER: l'infrastruttura e i servizi di calcolo e archiviazione distribuiti per le comunità scientifiche e non in Emilia Romagna



Marco Bencivenni¹, Daniele Cesini¹, Alessandro Venturini²
(CNR/ISOF), Paolo Veronesi¹

¹INFN-CNAF, ²CNR-ISOF, Istituto per la Sintesi Organica e la Fotoreattività

Abstract. 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 di ricerca. Questi paradigmi di calcolo e di accesso ai dati hanno dato un grande impulso alla ricerca producendo importanti risultati in brevissimo tempo. Rete ad alta velocità e certificata 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. È 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, aperta agli istituti e alle comunità scientifiche esistenti sul territorio della regione Emilia Romagna.

1. Introduzione

Le tecnologie del calcolo distribuito, quali la Grid [1] e più recentemente anche il Cloud Computing [7], hanno dimostrato vantaggi non solo in termini economici, ma soprattutto dal punto di vista della promozione di collaborazioni e dello scambio scientifico fra comunità diverse. La grande quantità di dati prodotti dalla ricerca nei campi della Fisica delle Alte Energie, astronomia e 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 nuove aree di interesse, quali la meta-genomica, sta suggerendo che l'integrazione fra discipline differenti può contri-

buire ad affrontare problemi ancora senza risposta. È altresì evidente che la condivisione di middleware e infrastrutture comuni stimola iniziative scientifiche multi-disciplinari, permette il trasferimento di tecnologie 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 nell'ambito del 6° (DataGRID [8], DataTAG [9]) e poi 7° programma quadro europeo (EGEE [10], EGI [11], PRACE [13]), 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) [3]. 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 ad esempio la sanità e la pubblica amministrazione.

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 l'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 GARR e possono inoltre usufruire di una serie di servizi base grazie all'appartenenza alla federazione di identità 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.

È quindi nata DUCK (Distributed Unified Computing for Knowledge) [2], una collaborazione 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

ti Climatici;

- Centro Interdipartimentale di Ricerca sul Cancro "Giorgio Prodi" (Università di Bologna);
- Biocomputing Group, Dipartimento di Biologia Evoluzionistica Sperimentale dell'Università di Bologna;
- CeIRG Centro Interdipartimentale di Ricerche Genomiche dell'Università di Modena e Reggio Emilia;
- Area della 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.

2. Le attività e l'infrastruttura di calcolo

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 cal-

- colocalizzato;
- 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 di DUCK, sono stati realizzati un sito web [2] e un'infrastruttura composta dalle strutture di ricerca delle organizzazioni partecipanti, illustrata in figura 1 (in viola i centri di calcolo in via di certificazione).

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

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 Organi-

zation (comput-er.it), abilitata nei diversi centri di calcolo, nella quale confluiscono le persone che partecipano alle diverse attività. Inoltre, sono stati configurati un catalogo dei dati (LFC) e un pool 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 livello sono gestiti dal nascente consorzio IGI (Italian Grid Infrastructure), nell'ambito delle attività di supporto e controllo dell'infrastruttura Grid italiana. Il personale di IGI inoltre collabora con i ricercatori degli altri istituti per analizzare le applicazioni proposte ed individuare gli scenari più opportuni per una loro esecuzione in ambiente Grid.

3. Un esempio di applicazione

A titolo di esempio si riporta l'esperienza realizzata nel campo della chimica computazionale, con un lavoro dedicato allo studio della dinamica di una membrana lipidica, in un sistema grande e complesso formato da 5783 molecole fra lipidi, molecole d'acqua e ioni, sotto l'effetto di un campo elettrico (in particolare vengono studiate due intensità di campo elettrico diverse) [3, 4]. L'applicazione si

basa sul Modelling Molecolare, che comprende tutti i metodi teorici e le tecniche computazionali utilizzate per rappresentare o simulare il comportamento delle molecole. I programmi di chimica computazionale permettono agli studiosi di generare dati molecolari quali le geome-

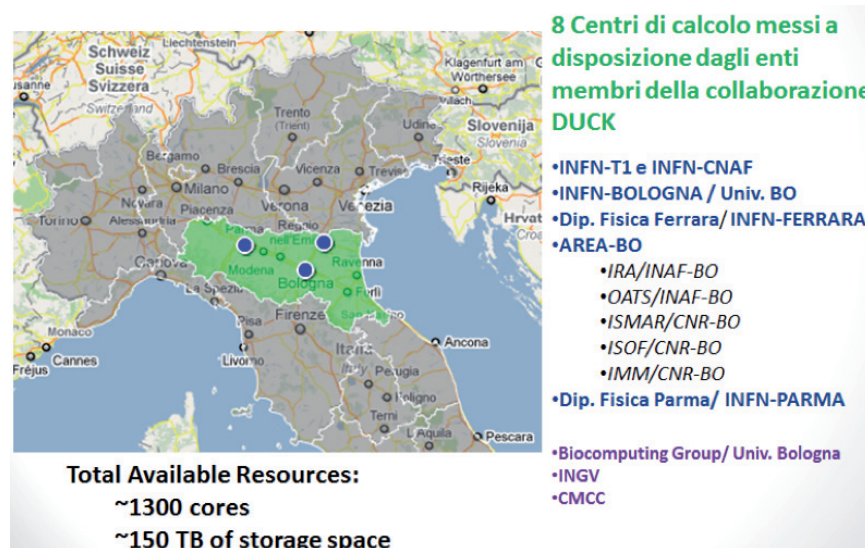


Fig. 1 L'infrastruttura Grid a supporto delle attività della collaborazione

<http://www.mcs.anl.gov/itf/Articles/WhatIsTheGrid.pdf>

[2] Comput-ER, <http://www.comput-er.it>

[3] IGI, Italian Grid Infrastructure
<http://www.italiangrid.it/>

[4] Alfieri R., Cesini D., Degli Esposti Boschi C., Ottani S., Salomoni D., Venturini A., “Da 20 anni nel futuro” Conferenza GARR 2011, Bologna, 8-10 Novembre 2011.

[5] Venturini A., F. Zerbetto, Phys. Chem. Chem. Phys., 2011, 13, 9216-9222.

[6] Phillips J. C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kale L., Schulten K., Scalable molecular dynamics with NAMD. J. Comp. Chem., 2005, 26, 1781-1802.

[7] A NIST Definition of Cloud Computing,
<http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>

[8] The DataGrid Project
<http://eu-datagrid.web.cern.ch/eu-datagrid>

[9] Datatag (Research & technological development for a Data TransAtlantic Grid)
<http://datatag.web.cern.ch/datatag>

[10] EGEE Portal: Enabling Grids for E-sciencE
<http://www.eu-egee.org>

[11] European Grid Infrastructure (EGI)
<http://www.egi.eu>

[12] EGI User Forum 2010, <http://uf2011.egi.eu>

[13] PRACE Research Infrastructure
<http://www.prace-ri.eu>

[14] Bartoli L., Montanucci L., Fronza R., Martelli P.L., Fariselli P., Carota L., Donvito G., Maggi G., Casadio R., J Proteome Res, 8:4362-4371, 2009.

[15] Piovesan D., Martelli P.L., Fariselli P., Zauli A, Rossi I, Casadio R. Nucleic Acids Res 39:W197-W202, 2011.

[16] Workshop DUCK, Bologna 10 novembre

2010, <https://agenda.italiangrid.org/conferenzeDisplay.py?confId=364>



Marco Bencivenni

comput-er@lists.infn.it

Laureato in Ingegneria delle Telecomunicazioni presso l'Università di Bologna. Lavora presso il CNAF, partecipa ai progetti EGI e IGI. Si occupa dello studio e implementazione di servizi

al fine di utilizzare le risorse Grid in modo facile e sicuro attraverso la realizzazione di un portale web.