

Telecontrollo basato su GRID e Cloud Computing

I. Bruno, N. Mitolo, P. Nesi

DISIT-DSI, Distributed Systems and Internet Technology Lab,
Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Firenze,
Firenze, Italy, <http://www.disit.dsi.unifi.it>
ivan.bruno@unifi.it, nicola.mitolo@unifi.it, paolo.nesi@unifi.it

Abstract

Nel controllo remoto di sistemi di misura, di sistemi domotici, di dispositivi fisici, la diagnosi e il controllo di qualità, l'organizzazione e l'analisi dei dati, l'ottimizzazione dei processi può risultare vantaggioso abbinare le tecnologie GRID e di Cloud computing per integrare e gestire da un unico punto di accesso sistemi eterogenei impegnati nelle fasi di acquisizione, elaborazione e cooperazione e per mettere a disposizione le risorse quando necessario. I principali contenuti innovativi del sistema proposto risiedono nell'ampia scalabilità della soluzione e nella modalità di definizione di processi GRID attraverso una descrizione formale delle risorse necessarie e le condizioni di schedulazione. Le risorse intese come nodi GRID (computer, server, cluster virtuali, storage di rete) sono individuate sulla base della loro profilazione (capacità computazionali, dimensioni dello spazio di storage) per effettuare l'allocazione e l'attivazione dei processi. La ricezione di notifiche via e-mail o sms (allarmi o semplice avviso di avvenuta esecuzione), e l'interrogazione via web permette di controllare lo stato di esecuzione dei processi e delle risorse e di intervenire in caso di malfunzionamenti. Tale tecnologia è attualmente utilizzata per il monitoraggio e il controllo remoto sullo stato di portali (accesso al portale, prestazioni di rete) nel settore dei social network e media service e dei loro sottosistemi (gestione failover e ripristino di processi e servizi). L'accesso ai servizi avviene attraverso un portale web semplice e intuitivo, nascondendo i dettagli implementativi legati all'utilizzo ottimale delle risorse di calcolo del GRID sottostante e consente l'accesso al sistema di controllo da sistemi mobili in ogni momento e in ogni luogo.

Introduzione

Le nuove tecnologie in ambito ICT stanno trasformando il modo di "comunicare" sulla rete in modo radicale e considerano l'informazione e la conoscenza come risorse strategiche. Il risultato è la produzione di grosse quantità di dati che necessitano di strategie organizzative in termini di risorse e un'elaborazione sempre più pesante in termini di analisi e utilizzo dell'informazione. Tali tecnologie da molti vengono definite con il neologismo di Pervasive Computing e tale dizione intende rappresentare un vasto insieme di elaboratori tra loro interconnessi: una rete capillare di micro e nano-computer, sensori biometrici e ambientali che, inseriti in ogni ambito della vita quotidiana, dall'abbigliamento alle abitazioni, dalle automobili al corpo umano, realizzano un ambiente simbiotico tra uomo e macchina totalmente connesso, una "rete di relazioni" che abbraccia dispositivi e persone. Di queste tecnologie, brevemente descritte di seguito, fanno parte i nuovi concetti di Internet of Things, Semantic Web e Big-Data.

L'Internet delle cose (*Internet of things*) è vista come una possibile evoluzione dell'uso della Rete. Gli oggetti si rendono riconoscibili e acquisiscono intelligenza grazie al fatto di poter comunicare dati su se stessi e accedere ad informazioni aggregate da parte di altri. Gli oggetti creano, quindi, un sistema pervasivo ed interconnesso avvalendosi di molteplici tecnologie di comunicazione (tipicamente a corto raggio). Il loro ruolo attivo grazie al collegamento alla rete consente un vasto campo di applicabilità: Domotica, Robotica, Avionica, Industria

automobilistica, Biomedicale, Monitoraggio in ambito industriale, Telemetria, Reti wireless di sensori, Sorveglianza, Rilevazione eventi avversi, Smart GRID e Smart City, Sistemi Embedded. E' tuttavia evidente la quantità di dati generata e da gestire sia in ottica di monitoraggio e controllo che di analisi e computazione per l'ottimizzazione delle prestazioni e dell'efficienza.

Nell'ambito del web semantico (*Semantic Web*) i documenti pubblicati (pagine HTML, file, immagini, e così via) sono associati ad informazioni e dati (metadati) che ne specificano il contesto semantico in un formato adatto all'interrogazione e l'interpretazione (es. tramite motori di ricerca) e, più in generale, all'elaborazione automatica e alla costruzione e gestione della conoscenza (ontologia). Con l'interpretazione del contenuto dei documenti che il Web semantico impone, è possibile effettuare ricerche molto più evolute delle attuali, basate sulla presenza nel documento di parole chiave e concetti, e altre operazioni specialistiche come la costruzione di reti di relazioni e connessioni tra documenti secondo logiche più elaborate del semplice collegamento ipertestuale.

Quando detto mette in evidenza come sempre più frequentemente si manifesti il problema della gestione di grandi aggregazioni di dati, la cui mole richiede strumenti differenti da quelli tradizionali, in tutte le fasi del processo (dalla gestione, al mantenimento, passando per condivisione, analisi e visualizzazione). Tali aggregazioni modellano e rappresentano l'interrelazione di dati provenienti potenzialmente da fonti eterogenee, quindi non soltanto i dati strutturati (come ad esempio quello dei database) ma anche non strutturati: immagini, email, dati GPS, informazioni derivate dai social network. L'insieme di tutti questi dati genera quel che si chiama Big Data consentendo a chi li analizza di ottenere una plusvalenza legata ad analisi più complete che consentono una conoscenza più fine degli utenti o degli oggetti, individuare nuove relazioni, definire nuove applicazioni e paradigmi. Le caratteristiche principali dei Big Data, si possono quindi riassumere nel: (i) volume dei dati gestiti, per ottenere analisi dettagliate e certe servono grandi quantità di dati da analizzare; (ii) varietà dell'informazione poichè oltre alla quantità, servono numerosi fonti diverse (strutturate e non); (iii) nella velocità di computazione per cui è essenziale effettuare analisi e ottenere dei dati in tempo reale; (iv) nel valore legato all'uso di modelli analitici complessi che impongono elevate capacità elaborative.

GRID e Cloud Computing

Le tecnologie emergenti e gli scenari descritti mettono in evidenza che la rete da sola non è sufficiente alla loro completa realizzazione. In quanto la complessità si sposta nell'organizzazione di grosse quantità di dati ed un'elaborazione sempre più pesante con un vincolo molto forte da rispettare: la riduzione dei tempi di risposta.

Il *GRID computing* combinato al *Cloud Computing* rappresenta l'evoluzione nell'utilizzo della rete Internet. Tramite queste due tecnologie, la rete Internet viene utilizzata per "condividere" potenza di calcolo e dati e consente di simulare un enorme computer virtuale che, in maniera intelligente, mette a disposizione le proprie risorse partendo da sistemi eterogenei connessi tra loro. In questo modo la potenza di calcolo viene messa a disposizione "on demand" ad un costo fino ad ora inaccessibile. I sistemi Cloud rappresentano un ambiente in cui le risorse di calcolo e di storage sono forniti come servizi a pagamento. Le risorse fisiche sono virtualizzate e condivise tra gli utenti, ciò garantisce la scalabilità delle risorse e delle capacità di calcolo e consente lo sviluppo di servizi di monitoraggio delle risorse cloud aggiuntive e di contabilizzazione del loro utilizzo. I sistemi di GRID

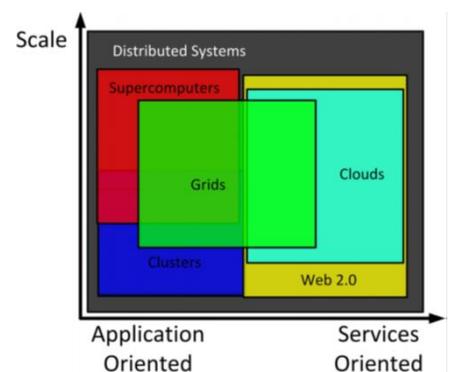


Figura 1 - Tratta da: "Cloud Computing and Grid Computing 360-Degree Compared" - Ian Foster et al. In Proc. of GCE'08, pp. 1-10, Nov. 2008.

computing permettono di definire applicazioni di calcolo ad elevate prestazioni e consentono di realizzare numerosi applicazioni in diversi ambienti. Il GRID computing consente un'elevata scalabilità, estendibilità, la distribuzione dell'elaborazione (Parallelismo), la selezione delle risorse di calcolo (nodi GRID) sulla base di profili computazionali (CPU, spazio disco, memoria, etc...), il controllo e monitoraggio puntuale delle risorse utilizzate (nodi GRID) e la schedulazione ottimizzata delle attività di elaborazione (Jobs). Quando le risorse Cloud vengono messe a disposizione e utilizzate come infrastrutture di calcolo per un ambiente GRID si può allora parlare di Cloud computing ([3]).

Scenari e contesti applicativi

Le tecnologie internet, GRID e Cloud descritte possono essere utilizzate e combinate in vari contesti applicativi.

Nel Controllo di Sistemi di Misura il sistema di gestione, monitoraggio a distanza di impianti, strumentazioni e processi di misura integra le fasi di acquisizione, elaborazione e cooperazione di sistemi non omogenei. Il problema nella gestione di sistemi non omogenei è legato principalmente ai protocolli di comunicazione. La gestione della comunicazione verso tali sistemi può essere superata se vengono utilizzate tecniche di controllo basate su GRID. Ogni sistema viene gestito autonomamente da un processo GRID il quale è in grado di realizzare un interfacciamento tra il sistema di controllo vero e proprio e il sistema controllato. In altre parole il processo GRID uniforma i sistemi controllati da un punto di vista del protocollo di comunicazione dell'intero sistema. Questa soluzione permette la scalabilità del sistema.

Nel Controllo di dispositivi (attuatori, rilevatori, sensori, etc...) tipicamente si conosce puntualmente lo stato di funzionamento, ma i dati rilevati, la loro contestualizzazione e le relazioni che su questi possono essere definite consentono la definizione e l'automazione di scelte comportamentali da parte del sistema di controllo. Ad esempio, nel caso di una gestione dei livelli d'illuminazione di ogni singolo punto luce dove occorre gestire la regolazione automatica in base alla programmazione impostata per fasce orarie e un controllo del funzionamento e segnalazione anomalie, l'analisi/elaborazione puntuale dei dati raccolti sulle condizioni ambientali dai sensori, rilevatori, etc., e la loro storicizzazione permette di realizzare un'ottimizzazione dei consumi e massimizzare il risparmio energetico. Tale soluzione di controllo mutuato dai sistemi domotici vede una realizzazione più ampia nella gestione del sistema di illuminazione di agglomerati urbani realizzati secondo il concetto delle *Smart City* ovvero delle città intelligenti.

Nella diagnosi e controllo di qualità, sono richiesti servizi di classificazione, previsione o analisi di correlazioni su dati provenienti da processi industriali e aziendali. La quantità di dati raccolta richiede oltre alla loro organizzazione, l'esecuzione di algoritmi computazionalmente pesanti e una riduzione dei tempi di risposta. L'elaborazione può essere realizzata ad sempio sfruttando la capacità di calcolo distribuita fornita dai sistemi GRID.

La Gestione e l'organizzazione dei processi organizzativi e comunicativi delle PA, richiede la gestione automatica dei documenti e l'interoperabilità tra PA. Sono quindi necessari meccanismi di condivisione dei dati/documenti, di ricerca documentale e di organizzazione sia su base keywords che su base semantica, di interazione e soluzioni di accessibilità all'informazioni sia da dispositivi consolidati come pc desktop che sistemi mobili di nuova generazione. Questo consente la definizione di nuovi servizi ai cittadini come quello dell'accesso in mobilità alle informazioni personali.

Per ognuno dei precedenti esempi si può prevedere un portale che consente un accesso ai servizi semplice e intuitivo, nascondendo i dettagli implementativi legati all'utilizzo ottimale delle risorse di calcolo del GRID sottostante. Se abbinato alle tecnologie basate su sistemi mobili, è possibile accedere al sistema di controllo in ogni

momento e in ogni luogo per monitoraggio, l'esecuzione di azioni remote e ricevere segnalazioni di allarmi o comunicazioni periodiche sullo stato di esecuzioni dei vari sistemi. Infine, mediante le soluzioni Cloud è possibile integrare tali sistemi realizzando un unico punto di accesso al GRID in grado di fornire al tempo stesso funzionalità di base e servizi atti a soddisfare le esigenze di sfruttamento delle risorse da parte delle aziende. Grazie ad una completa integrazione l'utente ha di fronte un unico punto di accesso ad un GRID polifunzionale che mette a disposizione servizi di base e applicazioni specifiche. Sia la gestione dei processi e dei file di dati, sia i risultati dell'elaborazione, sia le risorse disponibili e la configurazione generale del GRID, possono essere controllati da un'unica interfaccia web.

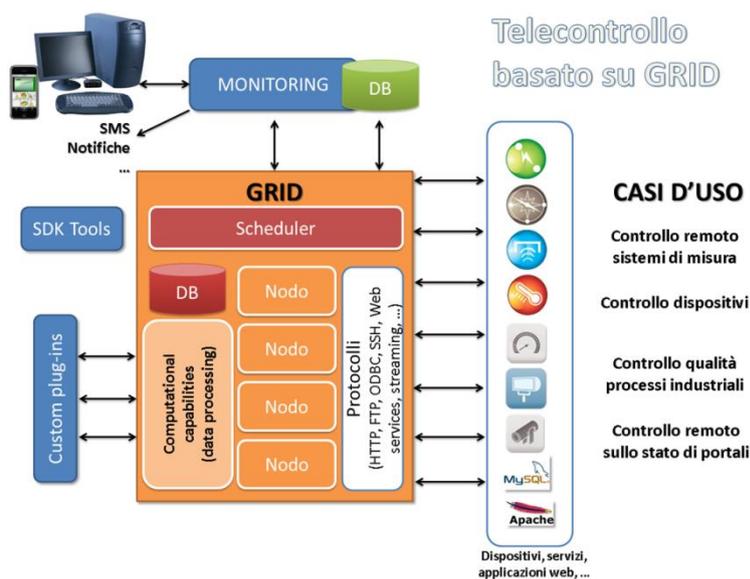


Figura 2 - Telecontrollo basato su GRID

Un soluzione GRID

Presso il DISIT Lab dell'Università di Firenze, è stato realizzato un sistema GRID e un ambiente di sviluppo di processi GRID ([1], [2]). Il sistema è in grado di gestire risorse di vario genere come: hosts, storage, network, databases, e canali di comunicazione esterni (protocolli http, ftp, Web Services, etc...). Esso è costituito da un *GRID Scheduler* in grado di effettuare: l'auto discovering dei nodi GRID, l'allocazione dei processi GRID selezionando dei nodi sulla base del profilo computazionale comunicato; il controllo dello stato nodi e dell'esecuzione dei processi; la collezione dei risultati, la gestione e la notifica errori. Un *Nodo GRID* è computer generico o dedicato industriale (virtuale o fisico), rappresenta il punto di accesso alle risorse computazionali, comunicazione, e di storage e comunica allo scheduler GRID lo stato esecuzione il profilo computazionale, i messaggi di errore, la disponibilità in termini di percentuale di utilizzo. Un *Request manager* realizza un front end verso il GRID che autorizza l'uso dei servizi e l'accesso, riceve e colleziona le richieste di esecuzione di processi GRID provenienti da: portali web, sistemi workflow, altri sistemi GRID, dai Nodi GRID interni o applicazioni IDE. Il sistema mette a disposizione un'applicazione IDE costituita da un Editor e debugger integrato per la formalizzazione e validazione dei processi GRID, l'installazione di processi GRID nello Scheduler e il lancio di processi nell'ambiente GRID. La definizione dei processi GRID presenta una descrizione testuale costituita da un insieme di metadati (autore, descrizione testuale, etc...) e una descrizione formale strutturata in informazioni di: (i) **Profilo** - risorse necessarie (spazio disco, banda, cpu...); (ii) **Schedulazione** - attivazione, periodicità, numero di tentativi in caso di failure, canale di notifica, ...; (iv) **Parametri** - elenco dei dati in ingresso e in uscita; (iv) **Funzionale** - Il codice di programmazione; (v) **Dipendenze** - elenco dei plug-ins utilizzati La programmazione dei processi GRID si basa sull'utilizzo di un linguaggio di scripting

(Javascript) opportunamente esteso con funzionalità in grado di gestire ad esempio differenti protocolli di comunicazione (http, ftp, web services wsdl, odbc, ssh, sftp), funzioni di sistema operativo, manipolazione xml ed una libreria SDK per lo sviluppo di plug-in accessibili direttamente dal linguaggio di scripting e quindi estendere dinamicamente il linguaggio stesso (Es. Funzionalità di Image Processing, Video Processing, Text Processing, etc...). L'utilizzo di una programmazione non vincolata ad un linguaggio compilato, consente il controllo e la modifica in tempo reale dell'esecuzione dei processi sia in fase di sviluppo che in esecuzione, con l'applicazione delle modifiche «a caldo».

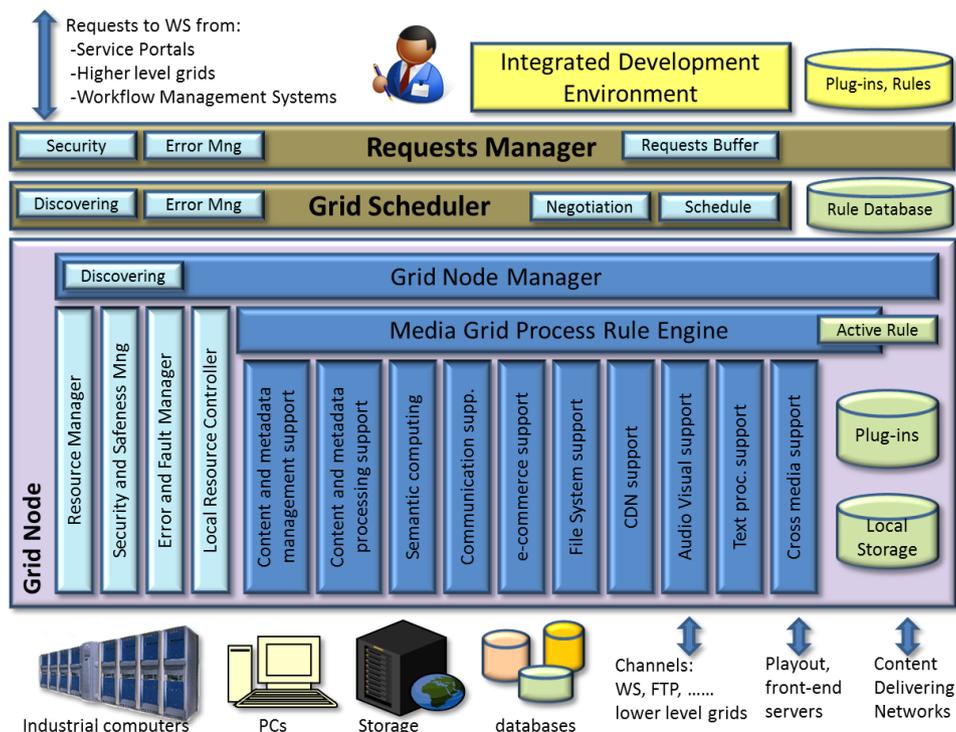


Figura 3 - Architettura di un sistema GRID

Ad oggi, tale tecnologia è stata utilizzata nel monitoraggio e il controllo remoto sullo stato di alcuni portali (accesso al portale, prestazioni di rete) nel settore delle smart city, social media service e dei loro sottosistemi (gestione failover e ripristino di processi e servizi). I portali non sono un semplice sito ma un cluster di macchine, processi e servizi in esecuzione su macchine virtuali all'interno di un datacenter (cloud). Le risorse impiegate comprendono: dei Web servers in balancing (per l'accesso alle pagine, l'upload di contenuti, le ricerche, accesso utente, ...); un Web Application server per i Servizi di storage dei metadata e contenuti, di indicizzazione e ricerca contenuti; un Database per l'organizzazione dei contenuti e metadata e la gestione dell'indicizzazione; un sistema di GRID computing per l'adattamento dei contenuti multimediali per la distribuzione multi-canale/multi-device, il massive Ingestion di contenuti, la Semantic Analysis e Computing, l'analisi dei metadata e la creazione delle relazioni tra contenuto/contenuto, utente/contenuto, utente/utente; una Storage Area costituita da dischi di rete.

Alcuni dei problemi di gestione che si riscontrano in tale configurazione sono: l'area di storage è piena o non raggiungibile, un web server è down, un servizio non risponde, la piattaforma GRID non risponde, ecc... Il funzionamento del portale quindi è legato al continuo e corretto funzionamento delle risorse utilizzate. Questo obiettivo è stato raggiunto realizzando un sistema di controllo e di monitoraggio basato su processi GRID definiti per: il monitoraggio di servizi (Web, porte...), processi (Web Server, Database, nodi GRID, Schedulers GRID, ...), risorse (consumo CPU, spazio disco, ...); realizzare un ontrollo periodico dello stato e riavvio dei servizi; calcolare

l'uptime; realizzare un sistema di alerting e notifica di anomalie via sms, mail o in push (es. servizio down, superamento soglie). Insieme al sistema GRID è stato realizzato un server di amministrazione, di collezionamento dati monitoraggio, un pannello di monitoraggio, e un'applicazione per l'accesso da smartphone al pannello di monitoraggio.

Conclusioni

Nel controllo remoto di sistemi di misura, di sistemi domotici, di dispositivi fisici, la diagnosi e il controllo di qualità, l'organizzazione e l'analisi dei dati, può risultare vantaggioso abbinare le tecnologie GRID e di Cloud computing per integrare e gestire da un unico punto di accesso sistemi eterogenei impegnati nelle fasi di acquisizione, elaborazione e cooperazione e per mettere a disposizione le risorse quando necessario. Il GRID computing consente inoltre di definire un controllo distribuito di risorse, dispositivi fisici e virtuali, attuatori, rilevatori, computer eterogenei e consente la gestione e l'analisi di grosse quantità di dati riducendo i costi in termini di tempo e la complessità dell'elaborazione. Infine è stata presentata e discussa una soluzione GRID e la sua applicazione in un contesto di monitoraggio e controllo di un portale di contenuti multimediali e sociali costituito da un sistema distribuito di risorse computazionali.

Riferimenti

1. Bellini P., Bruno I., & Nesi P. (2005). A distributed environment for automatic multimedia content production based on GRID. Proceedings - First International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution, AXMEDIS 2005, 2005 134-142.
2. Bellini P., Bruno I., Cenni D., Nesi P., (2012) "Micro GRIDs for scalable media computing and intelligence on distributed scenarios", IEEE Multimedia, IEEE Computer Soc. Press.
3. Foster I. et al. (2008) " Cloud Computing and GRID Computing 360-Degree Compared " - In Proc. of GCE'08, pp. 1-10, Nov. 2008.