

Cloud Computing in ENEA-GRID: Macchine Virtuali, Roaming Profile e Online Storage



Giovanni Ponti, Alessio Rocchi, Antonio Colavincenzo,
Gianfilippo Giannini, Alessandro Secco, Giovanni Bracco, Silvio
Migliori

ENEA – UTICT (Unità Tecnica per Information and Communication Technology)

Abstract. In questo lavoro, verrà descritta l'infrastruttura di cloud computing in ENEA-GRID che ha come obiettivo quello di fornire macchine virtuali per gli utenti di griglia. Tali macchine sfruttano e si poggiano sull'infrastruttura ENEA-GRID, la quale permette di attuare alcune soluzioni interessanti, quali la gestione dei profile degli utenti, l'integrazione delle macchine virtuali nella cella AFS ENEA e la possibilità di scambiare dati in modo semplice e sicuro utilizzando l'applicativo web OKBox sviluppato in ENEA. Sarà descritta anche l'apposita interfaccia per la gestione delle macchine virtuali integrata nel portale web FARO di accesso in ENEA-GRID e saranno discussi i primi risultati di stress test dell'infrastruttura di cloud.

1. Introduzione

L'ENEA, Agenzia Nazionale per le nuove tecnologie, l'energia e lo sviluppo economico e sostenibile, svolge attività di ricerca in diversi settori, quali efficienza energetica, fonti rinnovabili, fissione e la fusione nucleare, clima e ambiente e nuove tecnologie. Ha 14 sedi e centri di ricerca in territorio italiano, diversi uffici territoriali, un presidio di ricerca in Antartide e una sede di rappresentanza a Bruxelles, per un totale di oltre 2.600 dipendenti.

Per supportare le esigenze dei diversi gruppi di ricerca, è stata progettata e definita un'infrastruttura distribuita che prende il nome di ENEA-GRID [1]. Quest'ultima è nata nel 1998 e ha subito continue migliorie ed evoluzioni, che hanno portato oggi ad ottenere l'integrazione completa delle risorse computazionali, dei sistemi di storage, e dei sistemi di monitoring delle risorse presenti nei 6 centri di calcolo ENEA. Il tutto permette di offrire una capacità di storage integrata di circa 500 TB e una potenza di calcolo integrata di oltre 50 Tflops, la maggior parte della quale è erogata dagli oltre 6.000 core dei sistemi HPC CRESCO [2], che hanno nel Centro Ricerche ENEA di Portici il sito di maggiore

importanza.

ENEA-GRID offre una serie di risorse per supportare le attività scientifiche svolte dai ricercatori ENEA e dai loro collaboratori, sia per quanto riguarda applicazioni ad alto grado di parallelismo sia applicazioni seriali che richiedono un massivo utilizzo delle risorse. ENEA-GRID utilizza componenti software maturi, quali LSF per la gestione delle risorse, i due *file system* OpenAFS (geograficamente distribuito) a GPFS (per il calcolo parallelo), il sistema di autenticazione Kerberos 5, e il sistema di monitoring Zabbix. Inoltre, ENEA-GRID offre all'utente diversi modi di accesso, che vanno dalla console remota SSH a interfacce grafiche *user friendly*, quali il desktop remoto attraverso NX e il portale web FARO [3].

In questo lavoro vogliamo fornire una panoramica dell'esperienza di Cloud Computing in ENEA-GRID, descrivendo come tali funzionalità si sono evolute nel corso degli anni fino ad arrivare al contributo principale qui presentato, vale a dire l'offerta di macchine virtuali per gli utenti di ENEA-GRID attraverso la piattaforma di cloud OpenNebula [4]. Queste macchine virtuali sono on-demand, istanziabili su richiesta degli utenti da un set predefinito di template. La ge-

stione di tali macchine è fatta in maniera efficiente mediante il salvataggio delle personalizzazioni utente (*roaming profile*), e questo meccanismo permette di gestire le risorse in maniera efficiente in quanto, una volta terminato l'utilizzo, le macchine virtuali possono essere distrutte per liberare le risorse. Inoltre, lo scambio dei dati tra la macchina virtuale e la macchina fisica dell'utente avviene tramite un sistema di storage online implementato in ENEA che prende il nome di OKBox.

L'articolo è organizzato nel seguente modo: verrà fatta una panoramica delle funzionalità di cloud computing disponibili in modalità nativa in ENEA-GRID, per poi entrare più nel dettaglio su due diverse tipologie di sistemi di cloud presenti nella griglia. Saranno quindi descritti i sistemi di cloud per servizi virtualizzati, per poi proseguire con l'obiettivo principale di questo lavoro, ossia i servizi di cloud per erogare macchine virtuali per gli utenti. Per queste ultime, sarà descritto il modo in cui salvare le personalizzazioni utente (*roaming profile*) e come utilizzare il servizio di storage online per lo scambio dei dati (ENEA OKBox). Infine, saranno discussi i risultati di un primo stress test effettuato sull'infrastruttura.

2. Cloud computing in ENEA-GRID

ENEA-GRID esporta intrinsecamente e fin dalla sua nascita funzionalità oggi peculiari dei sistemi di cloud computing. In particolare, si fa riferimento allo storage distribuito messo a disposizione da OpenAFS, il quale permette di accedere ai propri dati in griglia in maniera indipendente dalla postazione dalla quale ci si trova. Dal punto di vista delle applicazioni, ENEA-GRID è nata per l'utilizzo di software remoti e, in questa direzione, ha subito delle evoluzioni che hanno portato alla recente definizione dei Laboratori Virtuali [5]: aree accessibili via web che offrono un ambiente di servizi di utilità per una particolare area tematica e supportano utenti che vogliono collaborare in alcuni scenari specifici, permettendo la condivisione di documenti, opinioni, e l'accesso a software dedicati.

Vediamo ora quali sono stati i passi successivi che hanno permesso di estendere le funzionalità di ENEA-GRID in senso *cloud-oriented*. In particolare, faremo riferimento alla virtualizzazione di servizi e alle macchine virtuali per gli utenti.

2.1 Cloud per servizi virtualizzati

I servizi ICT virtualizzati di ENEA hanno permesso di ampliare l'offerta di funzionalità per gli utenti. Tali servizi possono consistere nell'accesso ad applicativi specifici o ad interi sistemi customizzati secondo le richieste specifiche che vengono forniti da centri di calcolo di taglia importante, capaci di rilevanti economie di scala, attraverso in generale l'utilizzo di risorse di calcolo virtualizzate. Per erogare questi servizi sono utilizzati prodotti stabili e consolidati, che si basano sulla piattaforma VMware e garantiscono soluzioni efficienti e affidabili per servizi critici.

Questo sistema permette di far fronte a richieste specifiche di gruppi di lavoro e di esigenze di progetti di ricerca senza tuttavia stravolgere l'infrastruttura fisica di ENEA-GRID, sia dal punto di vista hardware che software. In questo modo, i due concetti di Grid Computing e di Cloud Computing possono integrarsi all'interno di un'infrastruttura come ENEA-GRID che, pur essendo distribuita geograficamente, presenta all'utente un ambiente unificato ed integrato.

2.2 Cloud per macchine virtuali utente

In questa sezione, sarà illustrato l'obiettivo principale di questo articolo, che consiste appunto nel descrivere tutte le attività che hanno portato ad erogare macchine virtuali per gli utenti di ENEA-GRID. In particolare, tutto è nato da una sperimentazione iniziata negli ultimi mesi del 2011, all'interno della quale è stata testata una prima installazione di OpenNebula all'interno di ENEA-GRID per erogare macchine virtuali [6].

Verificata la sostenibilità di questa sperimentazione sia dal punto di vista architetturale che software, il passo successivo ha portato alla definizione di una serie di template di macchine virtuali che gli utenti possono istanziare in base alle proprie necessità. Si noti che queste macchine virtuali sono "on demand" e, tipicamente,

hanno un ciclo di vita limitato alla singola sessione di utilizzo da parte dell'utente, il quale, una volta terminato il proprio task, può distruggerle (disposable VM) e rilasciare in questo modo le risorse occupate.

Tuttavia, sebbene tali macchine siano "non persistenti" (i.e., nessun cambiamento viene salvato sul disco virtuale), abbiamo pensato di adottare la gestione e il salvataggio delle personalizzazioni utente salvando il suo profilo (meglio noto in ambiente Windows come "roaming profile"). In questo modo, quando l'utente chiederà nuovamente di istanziare un particolare tipo di macchina virtuale, saranno contestualmente caricati il suo ambiente desktop e le sue precedenti personalizzazioni.

La gestione del profilo utente permette di definire dei template comuni per più utenti, e di avere una gestione efficiente e ottimizzata delle risorse fisiche dell'infrastruttura di cloud. Inoltre, questi profili vengono gestiti sia per macchine Windows che per macchine Linux. È opportuno sottolineare che tale gestione è stata resa possibile grazie all'integrazione delle macchine virtuali nella cella AFS di ENEA, che

permette di montare la home utente AFS nelle macchine virtuali e di accedere ai propri dati in ENEA-GRID. Per di più, l'integrazione in AFS ha evitato di creare utenze locali sulle macchine virtuali poiché l'autenticazione avviene in maniera integrata, basandosi sull'accesso ad una *active directory* per le macchine virtuali Windows e su Kerberos per quelle Linux, rendendo quindi i template ancora più generali e riutilizzabili per tutti gli utenti di griglia.

Per una migliore gestione dei template e delle risorse virtuali, sono state definite all'interno di OpenNebula delle apposite ACL, in modo tale che gli utenti possano utilizzare e gestire solo i template e le macchine virtuali che sono abilitati a coordinare. Inoltre, le ACL sono state definite anche sulle risorse fisiche, in modo tale da limitare, per ogni utente o gruppo di utenti, ad esempio il numero di macchine virtuali istanziate, la quantità di RAM occupata, il numero di CPU richieste, etc.

Per quanto riguarda l'accesso alle macchine virtuali, è stata implementata un'interfaccia grafica integrata all'interno del portale web FARO, illustrata nella seguente figura.

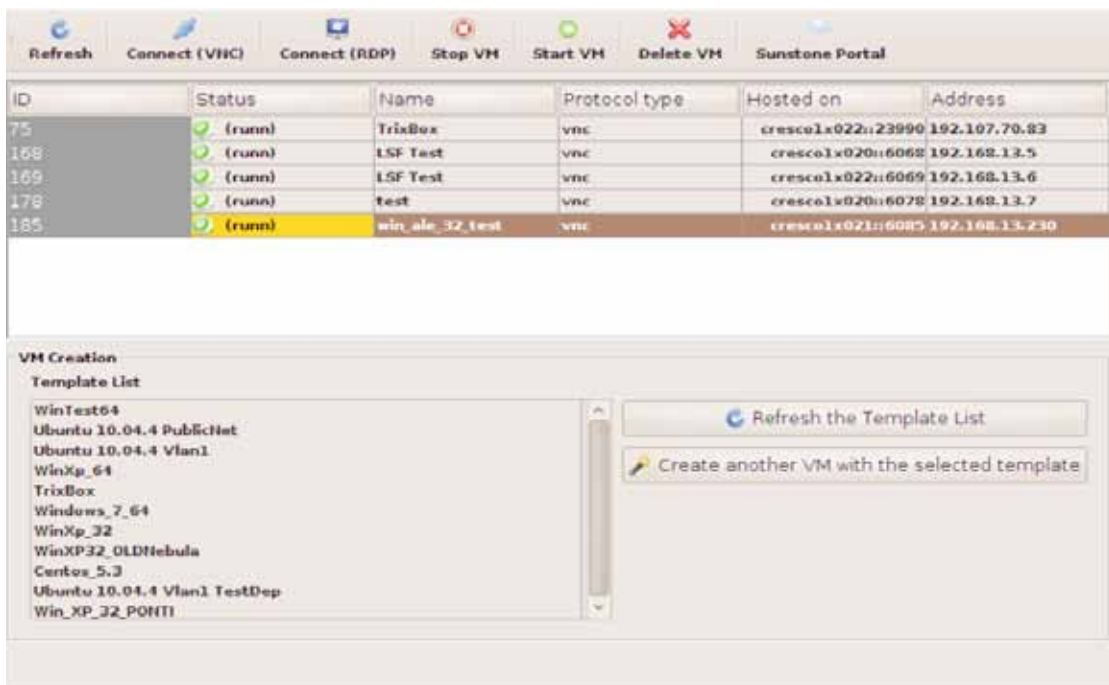


Fig 1 - FARO e OpenNebula – interfaccia di management in ENEA-GRID.

Si noti, in basso, la possibilità di selezionare il template da istanziare e, in alto, l'elenco delle macchine virtuali e i comandi per accedere alla macchina per controllarne l'esecuzione (start, stop, delete).

Si può notare come, anche in questo caso, ENEA-GRID offra una perfetta integrazione con le API di OpenNebula, permettendo la definizione di un'interfaccia grafica semplice, *user friendly*, e completamente funzionale. Questa interfaccia permette infatti di gestire le macchine virtuali e i template, fornendo un accesso immediato alle più comuni operazioni legate al ciclo di vita di una macchina virtuale, quali la creazione partendo da un template, lo *start*, lo *stop*, il *resuming*, e la *destroy*. Come si evince dalla figura, la lista dei template istanziabili può variare da utente a utente, in funzione delle ACL settate dagli amministratori; tuttavia, per ognuno dei template accessibili, l'utente è libero di scegliere quali risorse allocare e quante macchine creare e gestire (ovviamente, sempre con il vincolo delle ACL), senza doverne fare espressa richiesta agli amministratori. Inoltre, l'interfaccia è in grado di capire il tipo di macchina virtuale (ad esempio Windows o Linux) e, nel caso di macchina Windows, offre la possibilità di accedere, oltre che via VNC, anche via RDP, per migliorare l'efficienza.

Questa interfaccia esporta soltanto un sottoinsieme ristretto delle operazioni che possono essere eseguite sulla propria cloud. Per questo motivo, è stato predisposto anche l'accesso al portale web Sunstone di OpenNebula attraverso un apposito pulsante nell'interfaccia; in questo modo, gli utenti che vogliono gestire in maniera più fine le proprie risorse virtuali, avranno accesso ad una set completo di azioni.

3. Storage online e scambio dati con le macchine virtuali

Il contesto delle macchine virtuali presenta interessanti prospettive e si dimostra essere particolarmente versatile e interessante per soddisfare le richieste degli utenti senza stravolgere l'infrastruttura fisica di una rete di calcolo. Tuttavia, restano alcuni problemi pratici da gestire per rendere l'utilizzo di queste soluzioni semplice e immediato. Uno tra tutti riguarda il modo in cui i dati vengono trasferiti da e verso le macchine virtuali. In ENEA-GRID, la presenza del fi-

le system distribuito OpenAFS offre di per sé già un'ottima soluzione, permettendo di utilizzare la propria area utente per condividere file tra griglia, macchina client utente e macchine virtuali. Tuttavia, questo sistema prevede comunque l'installazione di un client e di meccanismi di autenticazione, e questa cosa può trovare alcuni limiti, specie se si pensa a dispositivi ubiqui quali smartphone e tablet.

Per coprire un *range* sempre più ampio di possibilità e fornire quindi un servizio sempre più completo, è stato implementato in ENEA un sistema di storage online accessibile via web e che fornisce una valida soluzione a questo problema. Nel seguente paragrafo descriveremo OKBox, un'applicazione web sviluppata in ENEA per supportare lo storage online.

3.1 ENEA OKBox

OKBox è un'applicazione web sviluppata in ENEA per lo storage online che implementa la politica "*Always and Anywhere on*" fornendo un portale web dal quale l'utente può caricare e scaricare i propri file, settare ACL e gestire la condivisione delle risorse avendo a disposizione soltanto la connettività in rete. L'archiviazione dei dati si basa su OpenAFS, mentre per l'autenticazione utilizza Kerberos.

Questo prodotto presenta notevoli vantaggi per l'utilizzatore, in quanto esporta una serie di funzionalità tutte integrate in un unico prodotto, e garantisce una gestione trasparente e sicura dei dati utente, poiché non si poggia su server di terze parti per archiviare i dati e non viola la *privacy* degli utenti. Inoltre, OKBox offre strumenti per favorire il lavoro collaborativo, permettendo di condividere dati con altri utenti, non necessariamente utenti di ENEA-GRID, semplicemente delegando e abilitando appositi parametri di accesso e condivisione.

Per tutta questa serie di motivi, OKBox è particolarmente utile se impiegato nel contesto delle macchine virtuali, in quanto gli utenti possono trasferire dati tra l'ambiente remoto virtuale e la propria macchina senza installare e configurare altri tool. Un esempio molto comune è la necessità di accedere, da una macchina vir-

tuale, alle stampanti locali installate sulla macchina fisica dell'utente; tramite OKBox è possibile quindi fare l'upload del file da stampare utilizzando il browser web della macchina virtuale e, contestualmente, accedere dalla macchina fisica (sempre tramite il browser web), recuperare il documento e stamparlo in locale.

4. Test dell'infrastruttura

In questo paragrafo, sarà discussa la fase di test dell'infrastruttura di cloud appena descritta per erogare macchine virtuali utente in ENEA-GRID. Nello specifico, si fa riferimento alla configurazione che prevede una cloud farm composta da 3 nodi IBM x3850/x3950-M2, ognuno equipaggiato con 4 CPU Xeon Quad-Core Tigerton E7330 (2.4GHz/1066MHz/6MB L2) e 32 Gb di RAM, per un totale di 48 core e 96 Gb di RAM come risorse fisiche. Inoltre, su uno di questi nodi girano anche i processi di OpenNebula, come da installazione standard consigliata dagli stessi sviluppatori.

L'obiettivo dello *stress test* è stato quello di verificare la responsività e l'usabilità del sistema in sotto diverse condizioni di carico. In particolare, si fa riferimento a due diverse fasi della sperimentazione, e cioè a una prima configurazione che non impiega tutte le risorse fisiche dell'infrastruttura e ad una seconda che sovraccarica l'infrastruttura allocando più macchine virtuali di quante le risorse fisiche sono capaci di sostenere. Entrambi i test sono stati fatti utilizzando un template basato su Windows XP 32bit SP3, che utilizza 1 CPU fisica e 1 Gb di RAM, e ogni macchina virtuale mantiene occupata la sua CPU eseguendo un test *benchmark* di fattorizzazione di grandi numeri interi.

I risultati di utilizzo dell'infrastruttura si sono rilevati ottimi nel primo caso (a sistema non completamente sovraccarico) non mostrando alcun rallentamento durante l'utilizzo delle macchine virtuali. Anche nel secondo caso (sistema completamente sovraccarico, come illustrato in Figura 2) il sistema ha mostrato una buona reat-

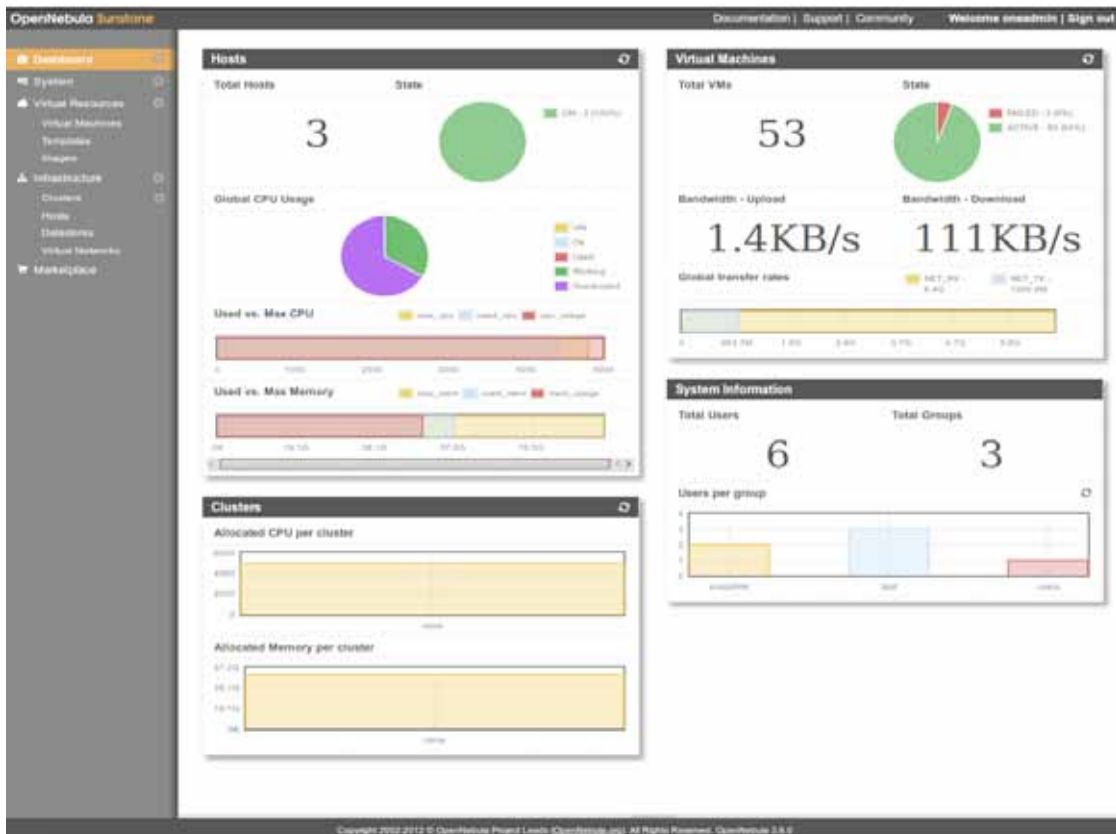


Fig 2 -Stress test dell'infrastruttura di cloud – Dashboard della Sunstone di OpenNebula.

tività, soprattutto nel caso in cui sulla macchina virtuale di test non si effettuano task impegnativi ma si simulano comportamenti standard e comuni, quali, e.g., la navigazione in internet e la video scrittura.

In Figura 2 è illustrata la *dashboard* della Sunstone di OpenNebula: In alto a sinistra, sono presentate le risorse fisiche occupate (numero di host, numero di CPU, e RAM); in alto a destra, le macchine virtuali istanziate e il loro stato; in basso, i cluster di macchine virtuali (a sinistra) e le informazioni su utenti e gruppi (a destra).

5. Conclusioni

In questo lavoro, è stata presentata l'infrastruttura di cloud computing sviluppata in ENEA-GRID. In particolare, è stata descritta la cloud basata su OpenNebula per fornire macchine virtuali per gli utenti ENEA. Queste macchine sono usa e getta, ma la gestione dei profili utente e la loro integrazione in AFS permette di conservare le personalizzazioni e di garantire l'accesso alle risorse della griglia. La flessibilità di OpenNebula ha permesso anche di sviluppare un'interfaccia grafica integrata nel portale web FARO di ENEA-GRID, attraverso la quale gli utenti possono accedere, gestire e amministrare le proprie risorse virtuali. Inoltre, è stato presentato OKBox come applicazione web capace supportare in modo semplice e sicuro la condivisione e lo scambio dei dati tra macchine client utente, griglia e macchine virtuali.

Da primi test effettuati sul sistema in diverse condizioni di carico, abbiamo verificato che l'infrastruttura permette di gestire le risorse in modo efficiente ed efficace, garantendo una responsabilità più che accettabile anche in condizioni di piena occupazione delle risorse.

Riferimenti bibliografici

- [1] ENEA-GRID – <http://www.eneagrid.enea.it/>
- [2] CRESCO – <http://www.cresco.enea.it/>
- [3] Rocchi A., Pierattini S., Bracco G., Migliori S., Beone F., Sciò C., Petricca A.: FARO: accesso WEB a risorse remote per l'industria e la ricerca.

Conferenza GARR, 2010

[4] OpenNebula – <http://www.opennebula.org>

[5] Laboratori Virtuali – <http://www.cresco.enea.it/virtulabs.html>

[6] Ponti G., Secco A., Ambrosino F., Bracco G., Ciavarella R., Colavincenzo A., D'Angelo P., De Rosa M., Funel A., Guarnieri G., Giammattei D., Migliori S., Pecoraro S., Petricca A., Pierattini S., Podda S., Rocchi A., Sciò C.: Esperimenti di Cloud Computing in ENEA-GRID. Conferenza GARR, 2011

Coautori e collaboratori

Hanno collaborato alla realizzazione di questo articolo: D. Abate, F. Ambrosino, G. Aprea, T. Bastianelli, F. Beone, M. Caporicci, M. Chinnici, A. Cucurullo, P. D'Angelo, A. Della Casa, M. De Rosa, A. Funel, G. Furini, D. Giammattei, S. Giuseppeponi, R. Guadagni, G. Guarnieri, A. Italiano, A. Mariano, G. Mencuccini, C. Mercuri, P. Orneli, S. Pecoraro, A. Perozziello, A. Petricca, S. Pierattini, S. Podda, F. Poggi, A. Quintiliani, C. Sciò, F. Simoni dell'ENEA – UTICT (Unità Tecnica per Information and Communication Technology).



Giovanni Ponti

giovanni.ponti@enea.it

È ricercatore ENEA dal 2010. Nel 2005 si è laureato con lode in Ingegneria Informatica presso l'Università della Calabria e, nel 2010, ha conseguito il

dottorato di ricerca in Ingegneria dei Sistemi e dell'Informazione. Le sue attività di ricerca riguardano i sistemi HPC, il Cloud Computing e il Data Mining. È coautore di articoli su riviste scientifiche internazionali, paper su atti di conferenza e capitoli di libro.