



PhD in INGEGNERIA DELL'INFORMAZIONE / INFORMATION TECHNOLOGY - 38th cycle

Research Area n. 1 - Computer Science and Engineering

PNRR_352 Research Field: DEVELOPMENT OF AN OPEN AND PORTABLE CLOUD MANAGEMENT FRAMEWORK FOR AUTOMATIC AND ADAPTIVE OPTIMIZATION OF APPLICATIONS BY MAPPING JOBS TO THE MOST SUITABLE RESOURCES IN A HETEROGENEOUS SYSTEM LANDSCAPE

Monthly net income of PhDscholarship (max 36 months)

€ 1250.0

In case of a change of the welfare rates during the three-year period, the amount could be modified.

Context of the research activity

Motivation and objectives of the research in this field

The worldwide public cloud market continues to grow and is expected to reach about 480 USD billions in 2022. Given the inherently distributed nature of cloud systems, coupled with an ever-increasing complexity in terms of heterogeneity of the supported applications, the enabling software stacks, and dynamicity of the cloud loads, guaranteeing the performance of mission-critical applications is extremely challenging.

In this context, Reinforcement Learning (RL) techniques have recently attracted increasing attention given the possibility of autonomously learning the optimal behaviour with little or even no prior knowledge of the system dynamics. By taking advantage of the possibility of distributed and/or hierarchical learning solutions, RL techniques have emerged as the most promising approach to cope in scenarios with loosely coupled components, running across several nodes which require fast, possibly local, decisions to provide consistent performance in face of uncertain and rapidly evolving environments.

This research will develop novel runtime management solutions based on RL enabling the execution of applications in cloud systems. The framework will allow



	<p>trading-off application performance (end-to-end latency and throughput) and energy consumption of the underlying infrastructure (in private cloud data centers)/cloud usage costs (in public cloud settings) while guaranteeing a reliable execution of the applications with performance guarantees.</p> <p>More specifically, the research will develop: a) framework to support the deployment of an application in cloud systems and to trigger application and resources reconfigurations, b) offline and online policy learning based on optimisation and RL techniques to adapt the system to evolving conditions of the environment or of the users' behaviour, c) validation of its solutions in industry settings.</p>
Methods and techniques that will be developed and used to carry out the research	<p>The techniques used and developed in this research fall under performance evaluation theory and applications, machine learning and reinforcement learning. The research will be based on machine learning approaches applied to predict the performance of cloud applications and optimization methods for the development of offline management policies. Several reinforcement learning methods will be investigated for online learning and runtime management. The final goal is to identify the minimum cost configuration of cloud applications which will provide performance guarantees (i.e., deadline for the execution of application requests/jobs) in private cloud data centers/public clouds.</p>
Educational objectives	<p>From an educational point of view, the research activity aims at teaching students cutting edge technology in use at cloud data centers and rigorous methods for the development of cloud resource selection and management systems that have strict performance and availability requirements.</p>
Job opportunities	<p>This research opens the doors to a career in the engineering of complex cloud systems supporting several applications domains ranging from, e.g., web service based applications, data-intensive and artificial intelligence applications with performance concerns. In addition to an academic career, the applicant will develop</p>



	expertise to support Small and Medium Enterprises or Public Administrations to: i) migrate their applications to the cloud, ii) identify the most suitable cloud provider/technology, and iii) optimise the application architecture to use at best the underlying cloud resources.
Composition of the research group	2 Full Professors 2 Associated Professors 2 Assistant Professors 2 PhD Students
Name of the research directors	Danilo Ardagna

Contacts	
Prof. Danilo Ardagna danilo.ardagna@polimi.it, 02 2399 3514, https://ardagna.faculty.polimi.it	

Additional support - Financial aid per PhD student per year (gross amount)	
Housing - Foreign Students	--
Housing - Out-of-town residents (more than 80Km out of Milano)	--

Scholarship Increase for a period abroad	
Amount monthly	625.0 €
By number of months	6

National Operational Program for Research and Innovation	
Company where the candidate will attend the stage (name and brief description)	E4 COMPUTER ENGINEERING SPA (https://www.e4company.com/)
By number of months at the company	6
Institution or company where the candidate will spend the period abroad (name and brief description)	E4 DEUTSCHLAND (https://www.e4company.com/)
By number of months abroad	6

Additional information: educational activity, teaching assistantship, computer availability, desk availability, any other information
<p>Attinenza alla tematiche, alle missioni/componenti del bando PNRR v. D.M. 352, art.6</p> <p>I sistemi cloud hanno un ruolo centrale per il PNRR. Le Amministrazioni sceglieranno se migrare verso una infrastruttura cloud nazionale (PSN) o verso soluzioni pubbliche a seconda della sensibilità dei dati e dei servizi coinvolti. Il cloud avrà un ruolo ed abilitante per molti servizi pubblici (Componente M1C1 - Migrazione al cloud delle amministrazioni centrali e locali, attraverso una infrastruttura nazionale).</p>



L'adozione di sistemi cloud favorirà anche l'innovazione e competitività del sistema produttivo (Componente M1C2) consentendo alle PMI italiane di accedere in modalità on demand a soluzioni infrastrutturali.

La ricerca svilupperà soluzioni innovative per la gestione dell'infrastruttura di un cloud data center che potenzialmente potranno essere adottate/fornire linee guida al PSN per garantire standard di prestazioni ed affidabilità in linea con gli investimenti: Investimento 1.2: Abilitazione e facilitazione migrazione al cloud per le PA locali, Investimento 1.4: Servizi digitali e cittadinanza digitale, Investimento 1.6: Digitalizzazione delle grandi amministrazioni centrali.

Impresa, presso cui si svolgerà l'attività esterna

E4 COMPUTER ENGINEERING è un Solution Provider riconosciuto a livello internazionale e coinvolto in molti progetti europei che sviluppa prodotti e servizi in ambito di High Performance Computing, High Performance Data Analytics, Intelligenza Artificiale e Deep Learning.

Il dottorando svilupperà e validerà presso E4 il sistema per il deployment e monitoraggio delle prestazioni delle applicazioni considerando hardware eterogeneo (ARM, x86, sistemi basati su GPU) e orchestratori per le risorse (Kubernetes, K3S, etc.). Inoltre verranno sviluppati e testati i meccanismi di adattamento dell'infrastruttura e di migrazione dei componenti software.

Ente, università, azienda, centro di ricerca presso cui si svolgerà il periodo di studio e ricerca all'estero.

E4 DEUTSCHLAND: L' entità tedesca è stata creata in forma di Niederlassung di E4 Computer Engineering SPA con sede a Monaco di Baviera.

E' un Solution Provider che progetta prodotti e servizi in ambito di High Performance Computing, High Performance Data Analytics, Intelligenza Artificiale e Deep Learning.

Il dottorando validerà presso la sede estera di E4 il framework sviluppato considerando diverse applicazioni di interesse e che tipicamente vengono eseguite su sistemi cloud.

All information regarding educational activities, personal funding, regulations and obligations of Ph.D. candidates are available on the web site <https://dottoratoit.deib.polimi.it/>