CV SUMMARY – DANILO ARDAGNA

DATE OF BIRTH 08/06/1973

EDUCATION

2004 **PhD Degree** in Computer Engineering, Politecnico di Milano, Italy.

2000 **Master of Science Degree** in Computer Engineering, Politecnico di Milano, Italy (Final Grade 100/100 cum laude).

CAREER

Dec 2014 – **Associate Professor**, Politecnico di Milano, Italy

2006 – 2014 **Assistant Professor**, Politecnico di Milano, Italy

2004 – 2006 **Post-Doc Researcher** (Assegnista di Ricerca), Politecnico di Milano, Italy

2001 – 2004 **Ph.D. Student,** Politecnico di Milano, Italy

HABILITATION

Type of habilitation	Country	SSD (if Italian habilitation) or topic area	Date of achievement
Full Professor	Italy	ING-INF/05	26/07/2018

RESEARCH INTERESTS

- **Big Data and Artificial Intelligence:** performance modelling, runtime resource allocation, scheduling, network architecture search, model optimal partitioning.
- **Cloud, edge computing, and Green IT:** design and runtime management, performance and energy trade-off, system optimization and game theory.
- **Service Oriented Architectures:** model driven development, Web services composition.
- **Autonomic Computing:** self-management techniques, dynamic resource allocation policies.
- **Design of IT Architectures:** cost analysis of IT architectures, system optimization.

LEADERSHIP IN COMPETITIVE RESEARCH PROJECTS

Total funding obtained from research projects in the last 10 years for my Institution about **2.95Mln**.

Project Acronym	**Time Period**	**Funding Institution**	**Funding Scheme**	**Role of the applicant**	**Budget for the applicant's institution**
AI-SPRINT	2021-	EU Commission	H2020 RIA	Project Coordinator	1.05Mln
ANDREAS	2021-2022	EU Commission	H2020 – TETRAMAX technology transfer	Responsible for POLIMI research unit	25k
ATMOSPHERE	2017-2019	EU Commission	H2020 RIA – Europe-Brazil collaboration	Responsible for POLIMI research unit	180k
DOSSIER-Cloud	2016-2018	EU Commission	H2020 Twinning	Participant – Proposal co-editor	140k
EUBRA-BIGSEA	2016-2017	EU Commission	H2020 RIA – Europe-Brazil collaboration	Responsible for POLIMI research unit	240k
DICE	2015-2018	EU Commission	H2020 RIA	Responsible for POLIMI research unit	530k
MODAClouds	2012-2015	EU Commission	FP7 - IP	WP Leader – Proposal co-editor	780k

SCIENTIFIC PRODUCTION AND METRICS

- **Scientific Productivity: 129** publications (**140** entries on Scopus, **234** co-authors according to Scopus, **0** single-author papers):
  - Author/Co-author of **29** top-ranked **Q1** journal papers (including IEEE Transactions on Cloud Computing, IEEE Transactions on Software Engineering. IEEE Transactions on Services Computing, IEEE Transactions on Dependable and Secure Computing, IEEE Transactions on Control Systems Technology, Computer Networks, INFORMS Journal on Computing, European Journal of Operational Research**)**
  - Author/Co-author of **48** scientific publications on peer-reviewed conferences including **1** top-level **A++/A+** conference (WWW)
  - Inventor/Co-inventor of **1 patent** with IBM
- **Publication Impact:** Based on Google Scholar: h-index **35** citations **6201**
  Based on Scopus: h-index **27** citations **3618**

AWARDS AND RECOGNITIONS (SELECTION OF PAST 10-YEARS)

1. *D. Ardagna*, F. Filippini, F. Magugliani, M. Cicala, K. K. Materka, M. Riedl, P. Skrzypek. Best project among the third call for Value Chain Technology Transfer Projects of H2020 TETRAMAX. ANDREAS selected for its exceptional contribution and innovation in power and cost management of deep learning training workloads. January 2022
2. *D. Ardagna*, S. Bernardi, E. Gianniti, S. Karimian Aliabadi, D. Perez-Palacin, J. I. Requeno. Modeling Performance of Hadoop Applications: A Journey from Queueing Networks to Stochastic Well Formed Nets. 16th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2016) Proceedings. 614-629. Granada, Spain. 2016. **Best paper award**
3. Evangelinou, M. Ciavotta, G. Kousiouris, *D. Ardagna*. A Joint Benchmark-Analytic Approach For Design-Time Assessment of Multi-Cloud Applications. In Cloud Forward (CF2015) Proceedings. 67-77. Pisa, Italy. 2015. **Best collaboration paper award** (European projects track)
4. **Google Cloud** for Education May 2022
5. **Windows Azure Research Pass** October 2013. Joint work with Prof. Elisabetta Di Nitto, DEEP-SE Cloud initiative
6. **AWS in Education research grant** March 2014. Joint work with Prof. Elisabetta Di Nitto, DEEP-SE Cloud initiative
7. **AWS in Education research grant** March 2012. Joint work with Prof. Elisabetta Di Nitto and Prof. Gianpaolo Cugola, DEEP-SE Cloud initiative
8. **IBM SUR Grant** October 2016. Joint work with Prof. Cristina Silvano and Prof. Letizia Tanca. P8 based cluster for Big Data research
9. Politecnico di Milano **Fondi 5 per mille** 2013. Poli-RISPOSTA **project co-applicant** (Principal Investigator Prof. F. Ballio)

INVITED TALKS AND SEMINARS (SELECTION OF PAST 10-YEARS)

1. AI-SPRINT: Future challenges of the Cognitive Cloud. European Commission. Digital Autonomy in the Computing Continuum. HORIZON-CL4-2022-DATA-01-03 call presentation. Online, 11/11/2021
2. Performance Prediction of GPU-based Deep Learning Applications. Spanish Supercomputing Network 13th Users Conference 2019, Zaragoza, Spain, 19/9/2019
3. Cloud and Multi-Cloud Computing: Current Challenges and Future Applications. A Quality of Service Perspective, 23/5/2015. PESOS Workshop at ICSE 2015
4. Model Driven Development and QoS Assessment of Cloud Applications, Polytechnic University of Catalonia, Barcelona, Spain, 7/5/2014
5. Model Driven Development and QoS Assessment of Cloud Applications, IBM T.J. Watson Research Center, Yorktown Heights, NY, USA, 15/10/2013
6. QoS Modeling and Evaluation in Cloud Environments, University of Austria at Timisoara, Timisoara, Romania, 24/9/2013
7. QoS Modeling and Evaluation in Cloud Environments, 7th Advanced Summer School in Service Oriented Computing, Crete, Greece, 4/7/2013
8. Generalized Nash Equilibria for the Service Provisioning Problem in Cloud Systems, IRISA, Rennes, France, 8/11/2012
9. Energy-Aware Autonomic Resource Allocation in Multitier Virtualized Environment (in very Large Service Centers), Imperial College, London, UK, 1/6/2012

TEACHING EXPERIENCE (SELECTION OF PAST 5-YEARS)

Institution name	Course name	Credits	No. of students	Reference Study Course	Time period	Students Evaluation
POLIMI	ALGORITHMS AND PARALLEL COMPUTING	10	274	Math Eng	2021-22	3.1/4
POLIMI	COMPUTING INFRASTRUCTURES	5	203	CSE	2021-22	-
POLIMI	ALGORITHMS AND PARALLEL COMPUTING	10	224	Math Eng	2020-21	3.2/4
POLIMI	COMPUTING INFRASTRUCTURES	5	303	CSE	2019-20	3.3/4
POLIMI	ALGORITHMS AND PARALLEL COMPUTING	10	182	Math Eng	2019-20	2.5/4
POLIMI	MULTIDISCIPLINARY PROJECT	5	31	CSE	2018-19	3.3/4
POLIMI	ALGORITHMS AND PARALLEL COMPUTING	10	154	Math Eng	2018-19	2.2/4
POLIMI	MULTIDISCIPLINARY PROJECT	5	24	CSE	2017-18	3.1/4
POLIMI	ALGORITHMS AND PARALLEL COMPUTING	10	127	Math Eng	2017-18	2.3/4
POLIMI	MULTIDISCIPLINARY PROJECT	5	11	CSE	2016-17	High
POLIMI	083559 - INFORMATICA B	7	188	Electrical Eng	2016-17	Medium
POLIMI	ALGORITHMS AND PARALLEL COMPUTING	10	129	Math Eng	2016-17	Medium

INSTITUTIONAL RESPONSIBILITIES (SELECTION OF PAST 10-YEARS)

2022 – present **Executive Board Member** of the Italian Computing and Data Infrastructure (ICDI) Working Group
2022 – present **Member** of the CINI Working Group on System and Service Quality
2022 – present **Member** of the National Research Center in High-Performance Computing, Big Data and Quantum Computing (Spoke Future HPC)
2020 – present **Member** of the Research Support Group of the Computer Science Research Area of the Department
2022 **Member** of the PhD Applications Evaluation Committee for the Computer Science Research Area
2015 – 2018 **Member** of IBM/POLIMI Collaborative Innovation Center on Big Data Analytics

SUPERVISION OF MASTER, DOCTORAL STUDENTS AND POSTDOCTORAL RESEARCHERS

2010 – present **Advisor/Co-advisor** of **4** Doctoral Students at Politecnico di Milano, and **7** Postdoctoral Researchers at Politecnico di Milano.
2012 – present **Opponent Member** of **7** Doctoral Examination Committees at Polytechnic University of Catalonia, University of Lille, INRIA-IRISA, University of Calgary, Università Tor Vergata

2001 – present **Advisor/Co-advisor** of **90+** Master Students in Computer Science Engineering, Politecnico di Milano, Italy

ORGANIZATION OF SCIENTIFIC MEETINGS (SELECTION OF PAST 10-YEARS)

2023 **Tutorials Co-chair**, SIGMETRICS 2023, Orlando, Florida, USA
2021 **General Chair**, IFIP WG 7.3 Performance 2021, Milan, Italy (virtual event)
2020 **General Chair**, IFIP WG 7.3 Performance 2020, Milan, Italy (virtual event)
2015 **General Chair**, ICGreen 2015, Milan, Italy

COMMISSIONS OF TRUST (SELECTION OF PAST 10-YEARS)

2022 **Expert Reviewer**, European Commission H2020 Center of Excellence project POP2 (Performance Optimisation and Productivity 2), 2022

2019 **Expert Reviewer**, SPEC (Standard Performance Evaluation Corporation) 2019 Award program committee

2019 **Expert Reviewer**, Barcelona Computing Center Stars program evaluator and **Chair** of the **evaluators** group, Spain
2017 **Expert Reviewer**, European Commission, ERC Consolidator grants
2016 **Expert Reviewer**, Netherlands Research Council for the Engineering and Applied Sciences Executive Board
2015 **Expert Reviewer**, Netherlands Research Council for the Engineering and Applied Sciences Executive Board
2013 **Expert Reviewer**, Netherlands Organisation for Scientific Research Executive Board
2013 **Expert Reviewer**, Austrian Science Fund Executive Board
2012 **Expert Reviewer**, Austrian Science Fund Executive Board

2018 – present **Associate Editor,** IEEE Transactions on Cloud Computing
2021 – present **Editorial Board Member,** KeAi-Elsevier (Open Publishing Program) International Journal of Intelligent Networks

2015 – 2020 **Associate Editor,** Springer open access BioMed Big Data Analytics journal

2014 – 2015, **General Chair** of the IEEE Special Technical Community on Sustainable Computing

2013 – 2014, **Secretary/Treasurer** of the IEEE Special Technical Community on Sustainable Computing
2011 – 2013, **Information Officer** of the IEEE Special Technical Community on Sustainable Computing

2017 – 2019, **Member** of AEIT (Associazione Italiana di Elettrotecnica, Elettronica, Automazione, Informatica e Telecomunicazioni) and AICT society

2011 – 2015, **Member** of the IT Systems Architect working group within AICA (Associazione Italiana per l'Informatica ed il Calcolo Automatico)

TWELVE MOST RELEVANT PUBLICATIONS

1. F. Filippini, M. Lattuada, M. Ciavotta, A. Jahani, *D. Ardagna*, E. Amaldi. A Path Relinking Method for the Joint Online Scheduling and Capacity Allocation of DL Training Workloads in GPU as a Service Systems. IEEE Transactions on Services Computing. To Appear. 1-16. DOI: 10.1109/TSC.2022.3188440. (Scimago Q1).

   Contribution: main publication of my PhD student's work. Led the problem definition and research agenda. Participated to the optimization problem formulation and validation.

   Impact: Reference paper of the TETRAMAX ANDREAS technology transfer project and TETRAMAX award, proposes a joint scheduler and resource allocator for deep learning jobs. Provides 95% costs savings in comparison to an IBM Research prototype.

2. M. Ciavotta, G. P. Gibilisco, *D. Ardagna*, E. Di Nitto, M. Lattuada, M. A. Almeida da Silva. Architectural Design of Cloud Applications: a Performance-aware Cost Minimization Approach. IEEE Transactions on Cloud Computing. To Appear. 1-18. DOI: 10.1109/TCC.2020.3015703. (Scimago Q1).

   Contribution: main publication of my PhD student's work. Led the research agenda, the optimization problem formulation and validation.

   Impact: Applied to an industry use case, provides between 21% and 85% costs savings in comparison to a SOTA alternative method.

3. M. Lattuada, E. Barbierato, E. Gianniti, D. Ardagna. Optimal Resource Allocation of Cloud-Based Spark Applications. IEEE Transactions on Cloud Computing. 10(2), 1301-1316. 2022. (Scimago Q1).

   Contribution: led the research agenda, the optimization problem formulation, the design of the heuristic method, and validation. Global optimum solution gap reduced by 30% with respect to a SOTA method by Spark inventors.

   Impact: global optimum solution gap reduced by 30% with respect to a SOTA method by Spark inventors.

4. E. Gianniti, M. Ciavotta, *D. Ardagna*. Optimizing Quality-Aware Big Data Applications in the Cloud. IEEE Transactions on Cloud Computing. 9(2), 737-752. 2021. (Scimago Q1).

   Contribution: main publication of my PhD student's work. Led the research agenda. Worked on the performance models and participated to the optimization problem formulation.

Impact: global optimum solution gap reduced by 10-30% with respect to a SOTA method by Spark inventors.

5. *D. Ardagna*, M. Ciavotta, R. Lancellotti, M. Guerriero. A Hierarchical Receding Horizon Algorithm for QoS-driven control of Multi-IaaS Applications. IEEE Transactions on Cloud Computing. 9(2), 418 - 434. 2021. (Scimago Q1).

   Contribution: Led the research agenda and the validation activities (cloud experiments and simulation). Participated to the optimization problems receding-horizon formulation.

   Impact: one of the few literature works providing a multi-time scale solution for resource management. With respect to heuristic provided by previous research works cost-savings range in 50-100%.

6. A. Maros, F. Murai, A. P. Couto da Silva, J. M. Almeida, M. Lattuada, E. Gianniti, M. Hosseini, *D. Ardagna.* Machine Learning for Performance Prediction of Spark Cloud Applications. IEEE Cloud 2019 Proceedings. 99-106. Milan, Italy. **Acceptance rate 20.8%**. (GGS A-).

   Contribution: responsible for the modelling of the computing/cloud aspects of the paper. Co-led the research agenda with Prof. Almeida an led the validation activities.

   Impact: Significantly improved the accuracy of Spark inventors' performance models reducing the percentage error from 126-187% to only 5-19% especially under resource contention and performance variability.

7. *D. Ardagna*, M. Ciavotta, M. Passacantando. Generalized Nash Equilibria for the Service Provisioning Problem in Multi-Cloud Systems. IEEE Transactions on Services Computing. 10(3), 381-395. 2017. (Scimago Q1).

   Contribution: Led the research agenda and participated to the game formulation. Led validation activities setup.

   Impact: Extension of my work in [selected publication 9] to multi-cloud scenarios (79 GoogleScholar citations).

8. B. Addis, *D. Ardagna,* B. Panicucci, M. Squillante, L. Zhang. A Hierarchical Approach for the Resource Management of Very Large Cloud Platforms. IEEE Transactions on Dependable and Secure Computing. 10(5), 253-272, 2013. (Scimago Q1).

   Contribution: summary of my ten years of work on Green IT. Led the research agenda and the definition of the hierarchical multi-time scale optimization problem. Led validation activities setup.

   Impact: one of the very few literature works (104 GoogleScholar citations) implementing resource management solutions which provides both performance and availability guarantees.

9. *D. Ardagna*, B. Panicucci, M. Passacantando. Generalized Nash Equilibria for the Service Provisioning Problem in Cloud Systems. IEEE Transactions on Services Computing. 6(4), 429-442, 2013. (Scimago Q1).

   Contribution: Led research agenda. Participated to game model formulation. Developed the toolchain used for the validation.

   Impact: reformulation of my work in [selected publication 11] as a potential game, providing a fully distributed implementation. Compared to other SOTA solutions Price of Anarchy improved by 50-70% (139 GoogleScholar citations).

10. *D. Ardagna*, B. Panicucci, M. Trubian, L. Zhang. Energy-Aware Autonomic Resource Allocation in Multi-tier Virtualized Environments. IEEE Transactions on Services Computing. 5(1), 2-19, 2012. (Scimago Q1).

    Contribution: led research agenda. Participated to the optimization model formulation. Coordinated the development of the main tool used for the validation. Performed validation activities.

    Impact: one of the very first works (259 GoogleScholar citations) on energy-aware resource management of cloud data centers. Improvement on average by 45% of the solution provided by the IBM Tivoli middleware.

11. *D. Ardagna*, B. Panicucci, M. Passacantando A Game Theoretic Formulation of the Service Provisioning Problem in Cloud Systems. WWW 2011 Proceedings (20th International World Wide Web Conference). 177-186. Hyderabad, India. **Acceptance rate 12.3%**. (GGS A++).

    Contribution: One of the very first work applying game theory to the management of cloud systems. Led research agenda, developed the main tool used for the validation. Participated to the game model formulation.

    Impact: one of the very first works (179 GoogleScholar citations) which applied Game Theory to the management of cloud services identifying Generalized Nash Equilibria.

12. *D. Ardagna*, B. Pernici. Adaptive Service Composition in Flexible Processes. IEEE Transactions on Software Engineering. 33(6), 369-384, 2007. (Scimago Q1).

    Contribution: main output of my post-doc research. Worked on the optimization problem formulation and negotiation framework setting. Toolchains development and run of the experimental campaign.

    Impact: one of the most cited literature works (1205 Google Scholar citations) for solving Web Services composition problem with global quality of service constraints.

CURRICULUM VITAE ET STUDIORUM

Danilo Ardagna

Politecnico di Milano,
Dipartimento di Elettronica, Informazione e Bioingegneria
Via Golgi 42, 201333 Milano, Italy
Telephone: +39-02 2399 3514
e-mail: danilo.ardagna@polimi.it
Born: June 8, 1973—Monza (MB), Italy
Nationality: Italian

Highlights

129 peer-reviewed publications, including **50 journal** papers (29 in Scimago Q1 ranking), **16 IEEE/ACM Transactions** and **2** top-level A++/A+ conferences

Co-author of **79** scientific publications on **peer-reviewed conferences/workshops**
Co-author of **6 chapters** in **scientific books**
Co-author of **two textbooks** in Italian
Co-editor of **one scientific book**
Co-inventor of **one international patent**
16 Invited Talks/Seminars/Panels (2007-2021)
3 tutorials
2 best paper awards
Visiting researcher and intern at **IBM T.J. Watson Research**, **BCAM (Basque Center for Applied Mathematics)**, and **Federal University of Minas Gerais**, Brazil
Co-applicant for several **European and national funded projects**, totaling **2.95mln euros** of funding to date
Google Cloud for Education May 2022
IBM 2016 SUR grant, big data P8 cluster donation. Joint work with Prof. Cristina Silvano and Prof. Letizia Tanca
IBM 2010 faculty award for research on virtualized systems
Top Compsci University Azure Adoption research July 2018, July 2017, October 2016
Amazon AWS in Education research grant May 2014, March 2012, and December 2010. Joint work with Prof. Elisabetta Di Nitto and Prof. Gianpaolo Cugola
Windows Azure Research Pass October 2013. Joint work with Prof. Elisabetta Di Nitto
Politecnico di Milano Fondi **5 per mille 2013 Poli-RISPOSTA project co-applicant**
Politecnico di Milano Fondi **5 per mille 2009 GAME-IT Principal Investigator**
Based on **Google Scholar** (27/7/2022), my **h-index** is **35** and total number of **citations** is **6201** (**2293 since 2017**), with first published paper in 2002. My **top ranked paper** was published in 2007 collecting up to **1205 citations** (**Scopus h-index 27**, **citations 3618**)

Table of Contents

Academic positions

POLITECNICO DI MILANO, MILAN, ITALY

Dec 2014 -	Associate Professor at the Department of Electronics, Information, and Bioengineering
2006-2014	Assistant Professor (tenure-track) at the Department of Electronics, Information, and Bioengineering
2004-2006	Post Doctorate at the Department of Electronics, Information, and Bioengineering. Grant by MIUR to support post-doctoral studies
2001-2004	Ph.D. Student. Grant by MIUR to support Ph.D. studies

FULL PROFESSOR QUALIFICATION

July 2018	Italian national scientific qualification in 09/H1 ING-INF/05 scientific sector (date of achievement 26/7/2018)

Education

2001-2004	Politecnico di Milano, Ph.D. in Computer Science and Engineering with the thesis "A cost-oriented methodology for the design of information technology architectures," advisor Prof. C. Francalanci
1992-2000	Politecnico di Milano, Master in Computer Science and Engineering (100/100 magna cum laude), with the thesis "ISIDE: A software tool for the cost-oriented design of Information Systems," advisor Prof. G. Bracchi

Research interests

AREAS OF SPECIALIZATION

Design, prototype, and evaluation of resource management algorithms for Cloud/Edge/Fog systems supporting Artificial Intelligence, Big Data and Service based applications. Optimization and game theory methods applied to capacity planning and management of very large-scale data centers. Performance modelling of Artificial Intelligence and Big Data applications. Machine learning models for Data Intensive and Deep Learning applications performance prediction. Model Driven Design and management of multi-Cloud applications. Energy efficiency solutions for virtualized infrastructures. Cost-oriented design and capacity planning of distributed IT architectures.

BIG DATA AND ARTIFICIAL INTELLIGENCE APPLICATIONS, EDGE COMPUTING

- Advanced scheduling policies for Deep Learning applications training in GPU based clusters
- Performance modelling of GPGPU systems through machine learning
- Optimal partitioning of Artificial Intelligence models in cloud and edge/fog systems and runtime management across computing continua
- Network architecture search of Deep Learning models with accuracy and performance guarantees
- Runtime management of Big Data and computing intensive applications: optimization models for the development of efficient runtime management policies
- Big Data applications performance modelling: machine learning and hybrid machine learning techniques for gray box performance modelling, simulation models based on Stochastic Well Formed Petri Nets for MapReduce and Spark
- Model-driven-development of Data-Intensive Applications: methodology and tools for data-aware quality-driven development and optimization of Big Data applications

GREEN AND CLOUD COMPUTING

- Model-driven-development of Cloud Applications: MDE approach for design and runtime management of applications targeting multiple Clouds with Quality of Service guarantees
- Service Provisioning: Definition of capacity allocation and load redirect policies with response time and availability constraints in Infrastructure as a Service Systems through optimization and game theory
- Energy Aware Resource Allocation: Development of autonomic resource allocation techniques for the resource management of very large-scale virtualized Data Centers

SERVICE ORIENTED ARCHITECTURES

- Model-driven-development of SOAs: Model driven development of component-based systems in SOA and pervasive environments
- Web Service Composition: Definition and analysis of service selection policies in composed Flexible Processes with Quality of Service global constraints

AUTONOMIC COMPUTING

- Self-Management Techniques: Design of network controllers for Autonomic Computing Infrastructures. Implementation of resource allocation policies for the optimization of Service Provider profits associated with multiple class of Service Level Agreement acting at multiple time scales

DESIGN OF IT ARCHITECTURES

- Cost comparison of IT architectures: Empirical analysis of the technical and management costs of different IT architectures. Implementation of a design tool to support minimum-cost IT architectural design (including joint hardware and network communication costs)

Awards

BEST PAPERS

- *D. Ardagna*, S. Bernardi, E. Gianniti, S. Karimian Aliabadi, D. Perez-Palacin, J. I. Requeno. Modeling Performance of Hadoop Applications: A Journey from Queueing Networks to Stochastic Well Formed Nets. ICA3PP 2016 Proceedings. Granada, Spain.
- Evangelinou, M. Ciavotta, G. Kousiouris, *D. Ardagna*. A Joint Benchmark-Analytic Approach For Design-Time Assessment of Multi-Cloud Applications. In CF2015 Proceedings. Pisa, Italy.

EUROPEAN PROJECTS

- *D. Ardagna*, F. Filippini, F. Magugliani, M. Cicala, K. K. Materka, M. Riedl, P. Skrzypek. Best project among the third call for Value Chain Technology Transfer Projects of H2020 TETRAMAX. ANDREAS selected for its exceptional contribution and innovation in power and cost management of deep learning training workloads. January 2022

GOOGLE CLOUD SERVICES

- Google Cloud for Education. May 2022 (5k$)

MICROSOFT CLOUD SERVICES

- Top Compsci University Azure Adoption. October 2016 (20k$), July 2017 (20k$), July 2018 (15k$)

- Windows Azure Research Pass October 2013. Joint work with Prof. Elisabetta Di Nitto, DEEP-SE Cloud initiative (40k$)

AMAZON WEB SERVICES

- AWS in Education research grant March 2014. Joint work with Prof. Elisabetta Di Nitto, DEEP-SE Cloud initiative (5k$)
- AWS in Education research grant March 2012. Joint work with Prof. Elisabetta Di Nitto and Prof. Gianpaolo Cugola, DEEP-SE Cloud initiative (10k$)
- AWS in Education research grant December 2010. Joint work with Prof. Elisabetta Di Nitto, DEEP-SE Cloud initiative (10k$)

IBM

- IBM SUR Grant October 2016. Joint work with Prof. Cristina Silvano and Prof. Letizia Tanca. P8 based cluster for Big Data research (80k€)
- IBM Faculty Award 2010. The PARVIS project has been funded by IBM to develop self-managing techniques for virtualized systems (20k$)

POLITECNICO DI MILANO

- Fondi 5 per mille 2013. Poli-RISPOSTA **project co-applicant** (Principal Investigator Prof. F. Ballio). Poli-RISPOSTA is one of the 8 projects selected out of 50 proposals by Politecnico di Milano to foster research on society. The project will develop an emergency management system for hydro-geological disasters (65k€)
- Fondi 5 per mille 2009. GAME-IT project **Principal Investigator**. GAME-IT is one of the 8 projects selected out of 77 proposals by Politecnico di Milano to support junior researchers. The project aims at the development of green policies for virtualized data centers (60k€)

International scientific collaborations

VISITING RESEARCHER

2016	One week visit at IBM T.J. Watson Research Centre. Participating to research projects on performance analysis and management of Spark applications with Machine Learning and Hybrid Machine Learning techniques, within the Infrastructures Performance Analysis and System Optimization Group
2013	Two weeks visit at IBM T.J. Watson Research Centre. Participating to research projects on management of Map-reduce Cloud data centers and analysis of clock drifts and synchronization algorithms for multi-Cloud systems, within the Infrastructures Performance Analysis and System Optimization Group
2011	One month visit at BCAM (Basque Center for Applied Mathematics) working within the Analysis, Design and Performance Evaluation of Telecommunication Networks and Computer Systems group on the study of Generalized Nash Equilibria for Platform-as-a-Service Cloud Systems
2007	Two weeks visit at IBM T.J. Watson Research Centre. Participating in a research project on the self-optimization of Autonomic Computing within the Infrastructures Performance Analysis and System Optimization Group
2005	One week visit at Computer Science Department at the Federal University of Minas Gerais, Brazil. Participating in a research project on the management of services and Service Level Agreements in the Service Oriented Architecture, within the Performance Analysis and Modelling group led by Prof. Virgilio Almeida

INTERNSHIP AT IBM T.J. WATSON RESEARCH CENTER

2003 Six months internship at IBM T. J. Watson Research Center, Performance Analysis and System Optimization Group. Participated in a research project on maximization of Service Level Agreement revenues in Web Systems. Grant by IBM Research to support studies

KEYNOTE SPEECHES

- Cloud and Multi-Cloud Computing: Current Challenges and Future Applications. A Quality of Service Perspective, 23/5/2015. PESOS Workshop at ICSE 2015

INVITED SEMINARS

1. AI-SPRINT: Future challenges of the Cognitive Cloud. European Commission. Digital Autonomy in the Computing Continuum. HORIZON-CL4-2022-DATA-01-03 call presentation. Online, 11/11/2021
2. Performance Prediction of GPU-based Deep Learning Applications. Spanish Supercomputing Network 13th Users Conference 2019, Zaragoza, Spain, 19/9/2019
3. Model Driven Development and QoS Assessment of Cloud Applications, Polytechnic University of Catalonia, Barcelona, Spain, 7/5/2014
4. Model Driven Development and QoS Assessment of Cloud Applications, IBM T.J. Watson Research Center, Yorktown Heights, NY, USA, 15/10/2013
5. QoS Modeling and Evaluation in Cloud Environments, University of Austria at Timisoara, Timisoara, Romania, 24/9/2013
6. QoS Modeling and Evaluation in Cloud Environments, 7th Advanced Summer School in Service Oriented Computing, Crete, Greece, 4/7/2013
7. Generalized Nash Equilibria for the Service Provisioning Problem in Cloud Systems, IRISA, Rennes, France, 8/11/2012
8. Energy-Aware Autonomic Resource Allocation in Multitier Virtualized Environment (in very Large Service Centers), Imperial College, London, UK, 1/6/2012
9. A Game Theoretic Formulation of the Service Provisioning Problem in Cloud Systems, IBM T.J. Watson Research Center, Hawthorne, NY, USA, 1/7/2011
10. A Game Theoretic Formulation of the Service Provisioning Problem in Cloud Systems, BCAM, Bilbao, Spain, 27/5/2011
11. Identification of Linear Parameter Varying State Space Models for the Performance Management of Autonomic Systems, IBM T.J. Watson Research Center, Hawthorne, NY, USA, 23/7/2009
12. Green Active Management of Energy in IT systems, Karlsruhe University, Karlsruhe, Germany, 29/4/2009
13. Active Management of Energy in IT systems, Italy, Università di Modena e Reggio Emilia, Modena, Italy, 24/10/2008
14. Adaptive Service Composition in Flexible Processes, IBM T.J. Watson Research Center, Hawthorne, NY, USA, 4/5/2007
15. Capacity Management and Planning for Modern Virtualized IT Infrastructures, IBM T.J. Watson Research Center, Hawthorne, NY, USA, 30/4/2007

TUTORIALS

1. Model-Driven Management of Multi-Cloud Applications, UCC 2014. London, UK, 11/12/2014. Joint tutorial with Giuliano Casale and Nicolas Ferry
2. Model Driven Design of Cloud Applications with a-priori Quality of Service Guarantees. SYNASC MICAS 2014. Timisoara, Romania, 22-23/9/2014. Joint tutorial with Marcos Almeida, Nicolas Ferry, and Juan F. Perez

3. Model Driven Design of Cloud Applications with a-priori Quality of Service Guarantees. ASE 2014. Västerås, Sweden, 16/9/2014. Joint tutorial with Marcos Almeida, Giuliano Casale, and Nicolas Ferry

Service activities

EDITORIAL BOARDS

- Since March 2018, Associate Editor of the IEEE Transactions on Cloud Computing
- Since October 2021, Editorial Board Member of the KeAi-Elsevier (Open Publishing Program) International Journal of Intelligent Networks
- 2017-2020. Associate Editor of the Springer open access BioMed Big Data Analytics journal
- 2015 Guest co-Editor, Springer Computing, Sustainable Computing Systems and Applications special issue
- 2014 Guest co-Editor, ACM SIGMETRICS Performance Evaluation Review, Performance and Resource Management in Big Data Applications special issue
- 2013 Guest co-Editor, Springer Computing, Cloud Computing special issue
- 2012-2019 Editorial Board Member of the International Journal of Cloud-Computing and Super-Computing

CONFERENCE ORGANIZATION COMMITTEES

- SIGMETRICS 2023 Tutorials Co-chair, Orlando, Florida, USA
- IFIP WG 7.3 Performance 2021 General chair, Milan, Italy (virtual event)
- IFIP WG 7.3 Performance 2020 General chair, Milan, Italy (virtual event)
- ICGreen 2015 General chair, Milan, Italy
- BPM 2008 Local organization chair, Milan, Italy

INTERNATIONAL WORKSHOPS CO-ORGANIZER AND CO-CHAIR

- Fourth International Workshop on Scalable Deep Learning over Parallel And Distributed Infrastructures (SCADL) 2022, at IPDPS 2022
- Third International Workshop on Scalable Deep Learning over Parallel And Distributed Infrastructures (SCADL) 2021, at IPDPS 2021
- Second International Workshop on Scalable Deep Learning over Parallel And Distributed Infrastructures (SCADL) 2020, at IPDPS 2020
- Second International Workshop On Quality-Aware Devops (QUDOS) 2016, at ISSTA 2016
- First International Workshop On Quality-Aware Devops (QUDOS) 2015, at ESEC/FSE 2015
- First International Workshop on Multi-Cloud applications and federated clouds (Multi-Cloud) 2013, at ICPE 2013
- First International Workshop on Run-time mOdels for Self-managing Systems and Applications (ROSSA) 2009, at Valuetools 2009
- First International workshop on Quality of Self-healing Web Services (QSWS) 2008, at BPM 2008

STEERING COMMITTEES

- Steering Committee of the International Workshop On Quality-Aware Devops (QUDOS) 2017-

INTERNATIONAL JOURNALS REVIEWING SERVICES (main)

- IEEE Transactions on Software Engineering
- IEEE Transactions on Services Computing
- IEEE Transactions on Computers
- IEEE Transactions on Cloud Computing
- IEEE Transactions on Storage
- IEEE Transactions on Network and Service Management
- IEEE Systems Journal
- ACM Transactions on the Web
- ACM Transactions on Modeling and Performance Evaluation of Computing Systems
- ACM Transactions on Software Engineering and Methodology
- ACM Transactions on Autonomous and Adaptive Systems
- IEEE Internet Computing
- IEEE Transactions on System Men and Cybernetics, Part C
- Journal of Systems and Software, Elsevier
- Journal of Parallel and Distributed Computing, Elsevier
- European Journal of Operational Research, Elsevier
- Journal on Selected Areas in Communications
- Computers in Industry, Elsevier
- Information and Software Technology, Elsevier
- Journal of Cooperative Information Systems
- Journal of Information Technology, Kluwer

INTERNATIONAL CONFERENCE TECHNICAL PROGRAM COMMITTEE MEMBER (main)

- IEEE CLOUD 2021-2009
- IEEE Big Data 2016-2014, 2019-2017 Senior PC member
- IEEE/ACM CCGrid 2022
- IEEE ICWS 2012- 2009
- IEEE/ACM UCC 2018- 2020
- IEEE ICDCS 2017, 2012, 2022
- IEEE ICPADS 2012
- IEEE DSS 2018-2017
- CIoT 2016
- IEEE BigDataSE 2016
- IEEE IC2E 2015
- OBD 2015
- IEEE ICAC 2015-2014
- IFIP DAIS 2014
- IEEE ICA3PP 2022, 2018, 2016-2015
- Closer 2019-2017, 2015- 2013
- IEEE BDDS 2013
- IEEE MASCOTS 2014-2013
- IEEE CGC 2013
- IEEE HPCC 2013
- IEEE SmartData 2017
- LOD 2019
- QoSA 2011-2009
- FiCloud 2014
- CSC 2012

- MobiWIS 2011
- CGC 2011
- SEASS 2011
- Intensive 2009
- IEEE/INFORMS SOLI 2009
- IESA 2007-2006

EVALUATION OF EUROPEAN AND NATIONAL RESEARCH PROJECTS

- External Reviewer for the European Commission, H2020 Center of Excellence project POP2 (Performance Optimisation and Productivity 2), 2022
- Barcelona Supercomputing Center Stars program evaluator and chair of the evaluators group, 2019
- ERC Consolidator grants expert evaluator for the European Commission, 2017
- Member of the Netherlands Research Council for the Engineering and Applied Sciences Executive Board, 2015, 2016
- Member of the Netherlands Organisation for Scientific Research Executive Board, 2013
- Member of the Austrian Science Fund Executive Board, 2012, 2013

EXTERNAL REFEREE ACADEMIC PROMOTION

- External referee for the assessment of University of Calgary applications for Promotion to Professor, 2018

ROLES WITHIN IEEE COMPUTER SOCIETY

- May 2014 - July 2015, General Chair of the IEEE Special Technical Community on Sustainable Computing
- 2013-2014, Secretary/Treasurer of the IEEE Special Technical Community on Sustainable Computing
- 2011-2103, Information Officer of the IEEE Special Technical Community on Sustainable Computing

ROLES WITHIN COMPUTER SCIENCE ITALIAN ASSOCIATIONS

- 2017-2020 member of AEIT (Associazione Italiana di Elettrotecnica, Elettronica, Automazione, Informatica e Telecomunicazioni) and AICT society. Organization of yearly seminars at my Department
- Since 2011 member of the IT Systems Architect working group within AICA (Associazione Italiana per l'Informatica ed il Calcolo Automatico)

INSTITUTIONAL RESPONSIBILITIES

- Since July 2022 Executive Board member of the Italian Computing and Data Infrastructure (ICDI) Working Group
- Since July 2022 member of the CINI Working Group on System and Service Quality
- Since July 2022 member of the National Research Center in High-Performance Computing, Big Data and Quantum Computing (Spoke Future HPC)
- June 2022 – Apr 2023, member of the PhD Applications Evaluation Committee for the Computer Science Research Area
- Since 2020 member of the Research Support Group of the Computer Science Research Area of the Department

- 2015-2018 member of IBM/POLIMI Collaborative Innovation Center on Big Data Analytics

Teaching experience

UNDERGRADUATE TEACHING

- 2006-2017, Principles of Computer Systems (Informatica B), Politecnico di Milano
- 2008-11, Computer Science Project (Progetto Ingegneria Informatica), Politecnico di Milano
- 2003-04, Web based Information Systems (Sistemi Informativi Web), Politecnico di Milano
- 2003-2008: Information Systems Platforms (Sistemi Informativi), Politecnico di Milano

GRADUATE TEACHING

- 2016- : Algorithms and Parallel Computing, Politecnico di Milano
- 2021-2022: Computer Infrastructures, Politecnico di Milano
- 2019-2020: Computer Infrastructures, Politecnico di Milano
- 2018-2019: Complex Networks, PhD course Politecnico di Milano (with Prof. Ana Paulo Couto Silva)
- 2015-2016: Computer Systems, Politecnico di Milano
- 2012-2014, 2015-2019: Multi-disciplinary Project, Politecnico di Milano, Master and PhD program Politecnico di Milano
- 2011-12; 2013-14: Cloud Computing from a Software Engineering perspective, Master and PhD program Politecnico di Milano (with Prof. Elisabetta Di Nitto)
- 2007-2008: Autonomic Services, PhD course Politecnico di Milano (with Prof. Barbara Pernici)
- 2015-2016: Graduate seminars on Computer Infrastructures, Politecnico di Milano
- 2013-2014: Graduate seminars on Software Engineering, Politecnico di Milano
- 2001-2006: Graduate seminars on Information System platforms and security, Politecnico di Milano
- 2005-: Instructor in many professional master programs in the context of the Politecnico educational offer through Cefriel and MIP (main topics: Information Systems, Performance Evaluation, Virtualization and Cloud Computing, Big Data, Hadoop, Spark, Spark Machine Learning and Deep Learning)

RESEARCH SUPERVISION

- 2021- : Supervisor for PhD. Student Bruno Guindani
- 2020- : Supervisor for PhD. Student Federica Filippini
- 2020- : Postdoctoral research advisor for Dr. Hamta Sedghani
- 2017-2020: Postdoctoral research advisor for Dr. Marco Lattuada (R&D STMicroelectronics)
- 2018-2019: Postdoctoral research advisor for Dr. Eugenio Gianniti (Free2move eSolutions)
- 2017- 2018: Postdoctoral research advisor for Dr. Enrico Barbierato (RTD-A, Università Cattolica)
- 2016-2017: Postdoctoral research advisor for Dr. Athanasia Evangelinou (European Commission Officer, HPC research unit)
- 2012- 2016: Postdoctoral research advisor for Dr. Michele Ciavotta (RTD-B, Università Milano Bicocca)
- 2009-2010: Postdoctoral research advisor for Dr. Barbara Stefania Panicucci
- 2015- 2018: Supervisor for PhD. Student Eugenio Gianniti
- 2013- 2016: Supervisor for PhD. Student Giovanni Paolo Gibilisco (R&D Akams)

- 2010-2012: Co-Supervisor for PhD. Student Niccolò Maria Calcavecchia
- 2015-2016: Supervisor of the Minor research for PhD. Student Alessandro Maria Rizzi
- 2015: Supervisor of the Minor research for PhD. Student Lorela Cano
- 2013: Supervisor of the Minor research for PhD. Student Santo Lombardo
- 2012: Supervisor of the Minor research for PhD. Student Silvia Lovergine
- 2010: Supervisor of the Minor research for PhD. Student Chiara Sandionigi

HOSTED VISITING STUDENTS

- 2020: Hamta Sedghani, University of Tabriz, Iran
- 2019: Arezoo Jahani, University of Tabriz, Iran
- 2016: Ehsan Ataie, Sharif University of Technology, Iran
- 2016: Soroush Karimian-Aliabadi, Sharif University of Technology, Iran
- 2014: Marzieh Malekimajd, Sharif University of Technology, Iran

PHD EXAMINER

- Yali Zhao, PhD Evaluator, The University of Melbourne, 2020
- SPEC 2019 Award program committee
- Marcelo Amaral, PhD Evaluation commission, Polytechnic University of Catalonia, 2019
- Gustavo Cipriano Motta Sousa, PhD Evaluation commission, University of Lille, 2018
- Yasaman Amannejad, PhD Evaluation commission, University of Calgary, 2017
- Marco Abundo, PhD Thesis reviewer, Università Tor Vergata, 2016
- Nicolas Poggi, PhD Evaluation commission, Polytechnic University of Catalonia, 2014
- Ajay Kattepur, PhD Evaluation commission, INRIA-IRISA, 2012

Participation in research projects

EUROPEAN RESEARCH PROJECT

2021- **Coordinator of** the AI-SPRINT H2020 project for the development of a design and runtime framework to support the design and operation of AI applications in computing continua with performance, security and privacy guarantees (**5mln euros** over 3 years, **1.1mln euros** allocated to my research unit).

2021- Participation to the LIGATE EURO-HPC research project at Politecnico di Milano. Develop performance models and Bayesian optimization methods for drug discovery application optimization in HPC clusters.

2020-2021 **Coordinator for Politecnico di Milano research unit** for the TETRAMAX ANDREAS project involved in the design of GPU scheduling policies for deep learning applications (**75k euros** over 10 months, **25k euros** allocated to my research unit)

2017-2019 **Coordinator for Politecnico di Milano research unit** for the ATMOSPHERE H2020 project involved in the performance modelling and prediction of container based and GPU based applications (**1.5mln euros** over 2 years, **180k euros** allocated to my research unit)

2016-2018 Participated to DOSSIER-Cloud Twinning project, providing expertise on Cloud services, resource and energy management for large scale data centers and big data systems. **Proposal co-editor** (**520k euros** over 3 years, **140k euros** allocated to my research unit)

2016-2017 **Coordinator for Politecnico di Milano research unit** for the EUBRA-BIGSEA H2020 project involved in the performance analysis and definition of run time management policies of Big Data and computing intensive applications (**1.5mln euros** over 2 years, **240k euros** allocated to my research unit)

2015-2018 **Coordinator for Politecnico di Milano research unit** for the DICE H2020 project on model driven design, quality assessment, and optimization of Data-Intensive applications.

Standardization leader and Ethics issue manager for data privacy (**4mln euros** over 3 years, **530k euros** allocated to my research unit)

2016-2017 **Coordinator for Politecnico di Milano research unit** for the UrbanMob LISA Cineca project on development of urban mobility applications with data and performance quality

2012-2015 Participated to MODAClouds IP research project at Politecnico di Milano on design and run-time management of applications targeting multiple Clouds. Developing design-time exploration with performance and availability guarantees. **WP Leader, proposal co-editor** (**9mln euros** over 3 years, **780k euros** allocated to my research unit)

2008-2013 Participation in SMScom IDEAS-ERC research project at Politecnico di Milano on design of situational applications with run-time performance guarantees

2008-2012 Participation in s-Cube, the European Network of Excellence in Software Services and Systems developing optimization algorithms for composed Web services

2008-2011 Participation in Q-ImPrESS FET/STREP research project at Politecnico di Milano on the model-driven-development of SOA systems with QoS guarantees

2005-2008 Participation in WS-Diamond FET/STREP research project at Politecnico di Milano on the design of repair actions in self-healing web services systems

ITALIAN NATIONAL RESEARCH PROJECT

2005-2008 Participation in DISCORSO FAR research project at Politecnico di Milano on the design of optimization and Web services composition algorithms to support Small and Medium Enterprises and industrial districts

2003-2005 Participation in MAIS FIRB research project at Politecnico di Milano on analysis of multi-channel adaptive Information Systems

Technology transfer

PATENTS

D. Ardagna, E. Conforti, C. Francalanci , M. Gatti, S. Lucchini, S. Morsello, M. Trubian. Method, system and computer program for configuring server farms at minimum cost. Politecnico di Milano - IBM. *Patent number US7886036*, USA Patent Office.

OPEN SOURCE TOOLS, PRODUCTS AND SERVICES

2020-2021 ANDREAS: Artificial intelligence traiNing scheDuler foR accElerAted resource clusterS. Advanced scheduler for GPU based clusters to support deep learning models training.
Funded by TETRAMAX technology transfer project, achieved TRL 8, commercialization already started (https://www.e4company.com/en/andreas-artificial-intelligence-training-scheduler-for-accelerated-resource-clusters/).

2021- SPACE4AI-D: System PerformAnce and Cost Evaluation on Cloud for AI applications Design. A novel open-source tool to support design space exploration, component placement and resource selection of AI applications running on cloud-edge systems.
One of the main assets of the on-going H2020 AI-SPRINT project (https://gitlab.polimi.it/ai-sprint/space4ai-d).

2017- aMLLib: a Machine Learning Library. An open-source library to train performance models for estimating the performance of complex software systems.
Applied to estimate performance of Spark systems and deep neural networks model training on GPUs. Allowed to improve the performance models of Spark inventors. Currently used within the H2020 AI-SPRINT and LIGATE research projects (https://github.com/a-MLLibrary/a-MLLibrary).

2016-2018 D-SPACE4Cloud: Data intensive System PerformAnce and Cost Evaluation on Cloud. DICE open-source tool for the design-time optimisation of big data applications based on Hadoop and Spark.

	One of the main H2020 DICE project assets led to the funding of the H2020 AI-SPRINT project and currently evolving into SPACE4AI-D (https://github.com/deib-polimi/diceH2020-space4cloudsWS).
2016-2017	OPT_IC & OPT_JR. A set of open-source run-time optimization-based resource management tools for Spark applications. One of the main H2020 EUBRA-BIGSEA project assets. Global optimum solution gap reduced by 30% with respect to an alternatice method by Spark inventors (https://github.com/eubr-bigsea/OPT_IC, https://github.com/eubr-bigsea/opt_jr).
2013-2015	SPACE4Cloud: System PerformAnce and Cost Evaluation on Cloud. An open-source design-time tool to support design space exploration of enterprise applications running on cloud. One of the main FP7 MODAClouds project assets. Led to the funding of the H2020 DICE project and evolved into D-SPACE4Cloud. provides between 21% and 85% costs savings in comparison with alternative methods (https://github.com/deib-polimi/modaclouds-space4cloud).
2006-2007	WASFO: Workload Analysis for Server Farm Optimization. A software tool for the capacity planning of autonomic virtualized server farms. One of the main outputs of my PhD studies, let to a patent already granted in USA and one pending patent in Europe.
2015-2016	Consultancy services: feasibility analysis for the migration of a virtual travel agency software to cloud for a leading operator in Italy.
2004-2005	Consultancy services: performance analysis of a virtualized Internet Banking software system within a sever consolidation project for a large bank in Italy.

General information on my open-source tools can be found at https://ardagna.faculty.polimi.it/?page_id=465

Full List of Publications

International Journals

1. F. Filippini, M. Lattuada, M. Ciavotta, A. Jahani, *D. Ardagna*, E. Amaldi. A Path Relinking Method for the Joint Online Scheduling and Capacity Allocation of DL Training Workloads in GPU as a Service Systems. IEEE Transactions on Services Computing. To Appear. 1-16. 10.1109/TSC.2022.3188440. (Scimago Q1).
2. S. Karimian-Aliabadi, M. M. Aseman-Manzar, R. Entezari-Maleki, *D. Ardagna*, B. Egger, A. Movaghar. Fixed-point Iteration Approach to Spark Scalable Performance Modeling and Evaluation. IEEE Transactions on Cloud Computing. To Appear. 1-14. DOI: 10.1109/TCC.2021.3119943. (Scimago Q1).
3. M. Ciavotta, G. P. Gibilisco, *D. Ardagna*, E. Di Nitto, M. Lattuada, M. A. Almeida da Silva. Architectural Design of Cloud Applications: a Performance-aware Cost Minimization Approach. IEEE Transactions on Cloud Computing. To Appear. 1-18. DOI: 10.1109/TCC.2020.3015703. (Scimago Q1).
4. E. Ataie, A. Evangelinou, E. Gianniti, *D. Ardagna*. A Hybrid Machine Learning Approach for Performance Modeling of Cloud-based Big Data Applications. The Computer Journal, Oxford Academic. 1-18. 10.1093/comjnl/bxab131. 2021. (Scimago Q2).
5. M. Lattuada, E. Barbierato, E. Gianniti, *D. Ardagna*. Optimal Resource Allocation of Cloud-Based Spark Applications. IEEE Transactions on Cloud Computing. 10(2), 1301-1316. 2022. (Scimago Q1).
6. M. Lattuada, E. Gianniti, *D. Ardagna*, L. Zhang. Performance Prediction of Deep Learning Applications Training in GPU as a Service Systems. Cluster Computing. 25, 1279-1302. 2022. (Scimago Q3).

7. H. Sedghani, *D. Ardagna*, M. Passacantando, M. Zolfy Lighvana, H. S.Aghdasi. An incentive mechanism based on a Stackelberg game for mobile crowdsensing systems with budget constraint. Ad Hoc Networks, Elsevier. 123, 1-15. 2021. (Scimago Q1).
8. *D. Ardagna*, E. Barbierato, E. Gianniti, M. Gribaudo, T.B.M. Pinto, A.P.C da Silva, J. M. Almeida. Predicting the Performance of Big Data Applications on the Cloud. Journal of Super Computing. 77(2), 1321-1353. 2021. (Scimago Q2).
9. E. Gianniti, M. Ciavotta, *D. Ardagna*. Optimizing Quality-Aware Big Data Applications in the Cloud. IEEE Transactions on Cloud Computing. 9(2), 737-752. 2021. (Scimago Q1).
10. *D. Ardagna*, M. Ciavotta, R. Lancellotti, M. Guerriero. A Hierarchical Receding Horizon Algorithm for QoS-driven control of Multi-IaaS Applications. IEEE Transactions on Cloud Computing. 9(2), 418 - 434. 2021. (Scimago Q1).
11. E. Ataie, R. Entezari-Maleki, L. Rashidi, K. S. Trivedi, *D. Ardagna*, A. Movaghar. Hierarchical Stochastic Models for Performance, Availability, and Power Consumption Analysis of IaaS Clouds. IEEE Transactions on Cloud Computing. 7(4), 1039-1056. 2019. (Scimago Q1).
12. S. Karimian-Aliabadi, *D. Ardagna*, R. Entezari-Maleki, E. Gianniti, A. Movaghar. Analytical Composite Performance Models for Big Data Applications. Journal of Network and Computer Applications. Elsevier. 142(15), 63-75. 2019. (Scimago Q1).
13. A. S Alic, J. Almeida, G. Aloisio, N. Andrade, N. Antunes, *D. Ardagna*, R. M. Badia, T. Basso, I. Blanquer, T. Braz, A. Brito, D. Elia, S. Fiore, D. Guedes, M. Lattuada, D. Lezzi, M. Maciel, W. Meira Jr., D. Mestre, R. Moraes, F. Morais, C. E. Pires, N. Puchalski Kozievitch, Walter dos Santos, P. Silva, M. Vieira. BIGSEA: A Big Data analytics platform for public transportation information. Future Generation Computer Systems. Elsevier. 96, 243-269. 2019. (Scimago Q1).
14. D. Ardagna, C. Cappiello, W. Samà, M. Vitali. Context-aware Data Quality Assessment for Big Data. Future Generation Computer Systems. Elsevier. 89, 548-562. 2018. (Scimago Q1).
15. L. Cano, G. Carello, D. Ardagna. A Framework for Joint Resource Allocation of MapReduce and Web Service Applications in a Shared Cloud Cluster. Journal of Parallel and Distributed Computing, Elsevier. 120, 127-147. 2018. (Scimago Q2).
16. M. Malekimajd, *D. Ardagna*, M. Ciavotta, E. Gianniti, M. Passacantando, A. M. Rizzi. Capacity Allocation and Admission Control of MapReduce Jobs. The Journal of Supercomputing. 74(10), 5314-5348. 2018. (Scimago Q2).
17. E. Ataie, R. Entezari-Maleki, S. E. Etesami, B. Egger, *D. Ardagna*, A. Movaghar. Power-aware Performance Analysis of Self-Adaptive Resource Management in IaaS Clouds. Future Generation Computer Systems. Elsevier. 86, 134-144. 2018. (Scimago Q1).
18. J. Anselmi, *D. Ardagna*, J. C.S. Lui, A. Wierman, Y. Xu, Z. Yang. The Economics of the Cloud. ACM Transactions on Modeling and Performance Evaluation of Computing Systems. 2(4), 1-23. 2017.
19. M. Scavuzzo, E. Di Nitto, *D. Ardagna*. Experiences and Challenges in Building a Data Intensive System for Data Migration. Empirical Software Engineering. 23(1), 52-86. 2018. (Scimago Q2).
20. A. Evangelinou, M. Ciavotta, *D. Ardagna*, A. Kopaneli, G. Kousiouris, T. Varvarigou. Enterprise Applications Cloud Rightsizing through a Joint Benchmarking and Optimization Approach. Future Generation Computer Systems. Elsevier. 78(1), 102-114. 2018. (Scimago Q1).
21. F. Ballio, D. Molinari, G. Minucci, M. Mazuran, C. Arias, S. Menoni, F. Atun, *D. Ardagna*, N. Berni, C. Pandolfo. The RISPOSTA procedure for the collection, storage and analysis of high quality, consistent and reliable damage data in the aftermath of floods. Journal of Flood Risk Management. Wiley on line. 604-615. 2018. (Scimago Q1).
22. E. Gianniti, A. M. Rizzi, E. Barbierato, M. Gribaudo, *D. Ardagna*. Fluid Petri Nets for the Performance Evaluation of MapReduce and Spark Applications. ACM SIGMETRICS Performance Evaluation Review. 44(4), 23-36. 2017.
23. *D. Ardagna*, M. Ciavotta, M. Passacantando. Generalized Nash Equilibria for the Service Provisioning Problem in Multi-Cloud Systems. IEEE Transactions on Services Computing. 10(3), 381-395. 2017. (Scimago Q1).

24. M. Ciavotta, G. P. Gibilisco, *D. Ardagna*. A Mixed Integer Linear Programming Optimization Approach for Multi-Cloud Capacity Allocation. Journal of Systems and Software. Elsevier. 123(1), 64-78. 2017. (Scimago Q1).
25. M. Passacantando, *D. Ardagna*, A. Savi. Service Provisioning Problem in Cloud and multi-Cloud Systems. INFORMS Journal on Computing. 28(2), 265-277. 2016. (Scimago Q1).
26. M. Malekimajd, *D. Ardagna*, M. Ciavotta, A. M. Rizzi, M. Passacantando. Optimal Map Reduce Job Capacity Allocation in Cloud Systems. ACM SIGMETRICS Performance Evaluation Review, 42 (4), 50-60. 2015.
27. *D. Ardagna*, G. Casale, M. Ciavotta, J. F. Perez, W. Wang. Quality-of-Service in Cloud Computing: Modeling Techniques and Their Applications. Journal of Internet Services and Applications. 5(1), 1-17, 2014. (Scimago Q2).
28. B. Addis, *D. Ardagna*, A. Capone, G. Carello. Energy-aware Joint Management of Networks and Cloud Infrastructures. Computer Networks. Elsevier. 70, 75-95, 2014. (Scimago Q1).
29. J. Anselmi, *D. Ardagna*, M. Passacantando. Generalized Nash Equilibria for SaaS/PaaS Clouds. European Journal of Operational Research. 236(1), 326-339, 2014. (Scimago Q1).
30. J. Anselmi, *D. Ardagna*, J. C. S. Lui, Adam Wierman, Y. Xu, Z. Yang. The economics of the cloud: price competition and congestion. ACM SIGMETRICS Performance Evaluation Review. 41(4), 47-49, 2014.
31. J. Anselmi, *D. Ardagna*, J. C. S. Lui, Adam Wierman, Y. Xu, Z. Yang. The economics of the cloud: price competition and congestion. SIGecom Exchanges 13(1), 58-63, 2014.
32. C. Sandionigi, *D. Ardagna*, G. Cugola, C. Ghezzi. Optimizing Service Selection and Allocation in Situational Computing Applications. IEEE Transactions on Services Computing. 6(3), 414-428, 2013. (Scimago Q1).
33. *D. Ardagna*, B. Panicucci, M. Passacantando. Generalized Nash Equilibria for the Service Provisioning Problem in Cloud Systems. IEEE Transactions on Services Computing. 6(4), 429-442, 2013. (Scimago Q1).
34. B. Addis, *D. Ardagna*, B. Panicucci, M. Squillante, L. Zhang. A Hierarchical Approach for the Resource Management of Very Large Cloud Platforms. IEEE Transactions on Dependable and Secure Computing. 10(5), 253-272, 2013. (Scimago Q1).
35. A. Koziolek, *D. Ardagna*, R. Mirandola. Hybrid Multi-Attribute QoS Optimization in Component Based Software Systems. Journal of Systems and Software. Journal of Systems and Software. Elsevier. 86(10), 2542-2558, 2013. (Scimago Q1).
36. *D. Ardagna*, B. Panicucci, M. Trubian, L. Zhang. Energy-Aware Autonomic Resource Allocation in Multi-tier Virtualized Environments. IEEE Transactions on Services Computing. 5(1), 2-19, 2012. (Scimago Q1).
37. *D. Ardagna*, S. Casolari, M. Colajanni, B. Panicucci. Dual Time-scale Distributed Capacity Allocation and Load Redirect Algorithms for Cloud Systems. Journal of Parallel and Distributed Computing, Elsevier. 72(6), 796-808, 2012. (Scimago Q2).
38. M. Tanelli, *D. Ardagna*, M. Lovera. Identification of LPV state space models for Autonomic Web service systems. IEEE Transactions on Control Systems Technology. 19(1), 93-103, 2011. (Scimago Q1).
39. *D. Ardagna*, L. Baresi, S. Comai, M. Comuzzi, B. Pernici. A Service-based Framework for Flexible Business Processes. IEEE Software. 28(2), 61-67, 2011. (Scimago Q2).
40. *D. Ardagna*, R. Mirandola. Per-flow Optimal Service Selection for Web Services Based Processes. Journal of Systems and Software, Elsevier. 83(8), 1512-1523, 2010. (Scimago Q1).
41. J. Almeida, V. Almeida, *D. Ardagna*, I. Cunha, C. Francalanci, M. Trubian. Joint Admission Control and Resource Allocation in Virtualized Servers. Journal of Parallel and Distributed Computing, Elsevier. 70(4), 344-362, 2010. (Scimago Q2).
42. *D. Ardagna*, C. Francalanci, M. Trubian. Joint Optimization of Hardware and Network costs for Distributed Computer Systems. IEEE Transactions on Systems Man and Cybernetics, Part A. 38, 470-484, 2008.
43. *D. Ardagna*, B. Pernici. Adaptive Service Composition in Flexible Processes. IEEE Transactions on Software Engineering. 33(6), 369-384, 2007. (Scimago Q1).
44. *D. Ardagna*, M. Comuzzi, E. Mussi, P. Plebani, B. Pernici. PAWS: a framework for processes with adaptive Web services. IEEE Software. 24(6), 39-46, 2007. (Scimago Q2).
45. *D. Ardagna*, M. Trubian, L. Zhang. SLA Based Resource Allocation Policies in Autonomic Environments. Journal of Parallel and Distributed Computing. 67, 259-270, Elsevier, 2007. (Scimago Q2).
46. *D. Ardagna*, B. Pernici. Dynamic Web Service Composition with QoS Constraints. International Journal of Business Process Integration and Management (IJBPIM). 1(4), 233-243, 2006. (Scimago Q3).
47. *D. Ardagna*, C. Francalanci, M. Trubian. A Multi-Model Algorithm for the Cost-Oriented Design of Internet-based Systems. Information Sciences, 176(21), 3105-3131, Elsevier. 2006. (Scimago Q1).

48. *D. Ardagna*, C. Francalanci. Joint Optimization of Hardware and Network Costs. Journal of Parallel and Distributed Computing 66(3), 469-485 Elsevier. 2006. (Scimago Q2).
49. *D. Ardagna*, C. Francalanci. A Cost-Oriented Approach for the Design of IT Architectures. Journal of Information Technology. 20(1), 32-51. 2005. (Scimago Q1).
50. *D. Ardagna*, C. Francalanci, V. Piuri. Designing and Rightsizing the Information System Architecture. Information Systems Frontiers 6(3), 229-245, Kluwer. 2004. (Scimago Q2).

International Conferences

1. A. Falanti, E. Lomurno, S. Samele, *D. Ardagna*, M. Matteucci. POPNASv2: An Efficient Multi-Objective Neural Architecture Search Technique. IEEE WCCI-IJCNN 2022 Proceedings (World Congress On Computational Intelligence, International Joint Conference on Neural Networks). 1-8. To Appear.
2. H. Sedghani, F. Filippini, D. Ardagna. A Random Greedy based Design Time Tool for AI Applications Component Placement and Resource Selection in Computing Continua. IEEE Edge 2021 Proceedings (2021 IEEE International Conference On Edge Computing). 32-40. Guangzhou, China (online). 2021.
3. H. Sedghani, F. Filippini, *D. Ardagna*. A randomized greedy method for AI applications component placement in Computing Continua. IEEE JCC 2021 Proceedings (12th IEEE International Conference On JointCloud Computing). Short paper. Online, 1-6. To Appear.
4. H. Sedghani, *D. Ardagna,* M. Matteucci, G. A. Fontana, G. Verticale, F. Amarilli, R. M. Badia, D. Lezzi, I. Blanquer, A. Martin, K. Wawruch. Advancing Design and Runtime Management of AI Applications with AI-SPRINT (Position Paper). COMPSAC 2021 Proceedings (IEEE 45th Annual Computers, Software, and Applications Conference). 1455-1462.
5. A. Gulino, A. Canakoglu, S. Ceri, *D. Ardagna*. Performance Prediction for Data-driven Workflows on Apache Spark. MASCOTS 2020 (28th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems). 1-8. Nice, France.
6. F. Filippini, M. Lattuada, A. Jahani, M. Ciavotta, *D. Ardagna*, E. Amaldi. Hierarchical Scheduling in on-demand GPU-as-a-Service Systems. SYNASC 2020 (22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing). 125-132. Timisoara, Romania.
7. A. Jahani, M. Lattuada, M. Ciavotta, *D. Ardagna*, E. Amaldi, L. Zhang. Optimizing on-demand GPUs in the Cloud for Deep Learning Applications Training. IEEE ICCCS 2019 Proceedings (4th IEEE International Conference on Computing, Communications and Security). 1-8. Rome, Italy.
8. A. Maros, F. Murai, A. P. Couto da Silva, J. M. Almeida, M. Lattuada, E. Gianniti, M. Hosseini, *D. Ardagna*. Machine Learning for Performance Prediction of Spark Cloud Applications. IEEE Cloud 2019 Proceedings. 99-106. Milan, Italy. **Acceptance rate 20.8%.** (GGS A-).
9. E. Gianniti, L. Zhang, *D. Ardagna*. Performance Prediction of GPU-based Deep Learning Application. Closer 2019 Proceedings. 279-286. Crete, Greece.
10. M. Lattuada, E. Gianniti, M. Hosseini, *D. Ardagna*, A. Maros, F. Murai, A. P. Couto da Silva, J. M. Almeida. Gray-Box Models for Performance Assessment of Spark Applications. Closer 2019 Proceedings. 609-618. Crete, Greece.
11. A. S. Alic, J. Almeida, W. Meira Jr., D. Guedes, W. dos Santos, I. Blanquer, S. Fiore, N. P. Kozievitch, N. Andrade, T. Braz, A. Brito, C. E. Pires, N. Antunes, M. Vieira, P. Silva, *D. Ardagna*, K. Fonseca, D. Lezzi, D. Elia, R. Moraes, T. Basso, W. H. Cavassin. GIS and Data: Three applications to enhance Mobility. Geoinfo 2018 Proceedings. 1-12. Campina Grande, Brazil.
12. E. Gianniti, L. Zhang, *D. Ardagna*. Performance Prediction of GPU-based Deep Learning Applications. SBAC-PAD 2018 Proceedings (30th International Symposium on Computer Architecture and High Performance Computing). 167-170. Lyon, France.
13. *D. Ardagna*, E. Barbierato, A. Evangelinou, E. Gianniti, M. Gribaudo, T. B. M. Pinto, A. Guimarães, A. P. Couto da Silva, J. M. Almeida. Performance Prediction of Cloud-Based Big Data Applications. ICPE 2018

Proceedings (9th ACM/SPEC International Conference on Performance Engineering). 192-199. Berlin, Germany.

14. M. Ciavotta, E. Gianniti, *D. Ardagna*. D-SPACE4Cloud: A Design Tool for Big Data Applications. ICA3PP 2016 Proceedings (16th International Conference on Algorithms and Architectures for Parallel Processing). 614-629. Granada, Spain.
15. *D. Ardagna*, S. Bernardi, E. Gianniti, S. Karimian Aliabadi, D. Perez-Palacin, J. I. Requeno. Modeling Performance of Hadoop Applications: A Journey from Queueing Networks to Stochastic Well Formed Nets. . ICA3PP 2016 Proceedings (16th International Conference on Algorithms and Architectures for Parallel Processing). 599-613. Granada, Spain. ***Best paper award***.
16. G. P. Gibilisco, M. Li, L. Zhang, *D. Ardagna*. Stage Aware Performance Modeling of DAG Based in Memory Analytic Platforms. In IEEE Cloud 2016 Proceedings. 188-195. San Francisco, USA. **Acceptance rate 15%.** (GGS A-).
17. M. A. Almeida da Silva, *D. Ardagna*, N. Ferry, Juan F.Perez. Constellation: A multi cloud application designed with the MODAClouds integrated modeling environment. In ICSSEA 2016 Proceedings (International Conference on Software & Systems Engineering and their Applications). Paris, France.
18. A. Evangelinoua, M. Ciavotta, G. Kousiouris, *D. Ardagna*. A Joint Benchmark-Analytic Approach For Design-Time Assessment of Multi-Cloud Applications. In CF2015 Proceedings (Cloud Forward Conference). 67-77. Pisa, Italy. ***Best collaboration paper award*** *(European projects track).*
19. *D. Ardagna*, G. P. Gibilisco, M. Ciavotta, A. Lavrentev. A Multi-Model Optimization Framework for the Model Driven Design of Cloud Applications. SSBSE 2014 Proceedings (6th International Symposium on Search-Based Software Engineering). 61-76. Fortaleza, Brazil.
20. D. Molinari, M. Mazuran, C. Arias, G. Minucci, F. Atun, *D. Ardagna*. Implementing tools to meet the floods directive requirements: a "procedure" to collect, store and manage damage data in the aftermath of flood events. FRIAR 2014 Proceedings (4th International Conference on Flood Recovery, Innovation and Response). 215-226. Poznan, Poland.
21. F. Giove, D. Longoni, M. Shokrolahi Yancheshmeh, *D. Ardagna*, E. Di Nitto. An approach for the Development of Portable Applications on PaaS Clouds. Closer 2013 Proceedings. 591-601. Aachen Germany.
22. *D. Ardagna*, S. Casolari, B. Panicucci Flexible Distributed Capacity Allocation and Load Redirect Algorithms for Cloud Systems. IEEE Cloud 2011 Proceedings. 163-170. Washignton DC, USA. **Acceptance rate 18%**. (GGS A-).
23. *D. Ardagna*, B. Panicucci, M. Passacantando A Game Theoretic Formulation of the Service Provisioning Problem in Cloud Systems. WWW 2011 Proceedings (20th International World Wide Web Conference). 177-186. Hyderabad, India. **Acceptance rate 12.3%.** (GGS A++).
24. *D. Ardagna*, C. Ghezzi, B. Panicucci, M. Trubian. Service Provisioning on the Cloud: Distributed Algorithms for Joint Capacity Allocation and Admission Control. Service Wave 2010 Proceedings. 1-12. Ghent, Belgium.
25. B. Addis, *D. Ardagna*, B. Panicucci, L. Zhang. Autonomic Management of Cloud Service Centers with Availability Guarantees. In IEEE Cloud 2010 Proceedings, 220-227. Miami, Florida, USA. **Acceptance rate 18%**. (GGS A-).
26. A. Martens, *D. Ardagna*, H. Koziolek, R. Mirandola, R. Reussner. A Hybrid Approach for Multi-Attribute QoS Optimisation in Component Based Software Systems. In QoSA 2010 Proceedings (6th International Conference on the Quality of Software Architectures). 84-101. Prague, Czech Republic.
27. *D. Ardagna*, M. Tanelli, M. Lovera, L. Zhang. Black-box Performance Models for Virtualized Web Service Applications. WOSP/SIPEW 2010 Proceedings (1st Joint WOSP/SIPEW International Conference on Performance Engineering). ACM DL, 153-164. San Jose, CA, USA.
28. M. Tanelli, N. Schiavoni, *D. Ardagna*, M. Lovera. Control-Oriented Multirate LPV Modelling of Virtualized Service Center Environments. CDC 2009 Proceedings (48th IEEE Conference on Decision and Control).
29. M. Tanelli, *D. Ardagna*, M. Lovera. LPV Model Identification in Virtualized Service Center Environments. SYSID 2009 Proceedings (15th IFAC Symposium on System Identification), 862-867, Saint-Malo, France.
30. M. Tanelli, *D. Ardagna*, M. Lovera. On- and off-line model identification for power management of Web service systems. CDC 2008 Proceedings (47th IEEE Conference on Decision and Control). 4497-4502, Cancun, Mexico.
31. M. Tanelli, *D. Ardagna*, M. Lovera, L. Zhang. Model Identification for Energy-aware Management of Web Service Systems. ICSOC 2008 Proceedings (6th International Conference on Service-Oriented Computing).

32. *D. Ardagna*, C. Cappiello, M. Lovera, B. Pernici, M. Tanelli. Active Energy-Aware Management of Business-Process Based Applications. Position Paper. Service Wave 2008 Proceedings, 183-195, Madrid, Spain, December 2008.
33. *D. Ardagna*, C. Ghezzi, R. Mirandola. Model Driven QoS Analyses of Composed Web Services. Service Wave 2008 Proceedings, 299-311, Madrid, Spain, December 2008.
34. *D. Ardagna*, C. Ghezzi, R. Mirandola. Rethinking the use of models in software architecture. QoSA 2008 Proceedings (4th International Conference on the Quality of Software Architectures). 1-27, Karlsruhe, Germany, October 2008.
35. B. Pernici, *D. Ardagna*, C. Cappiello. Business Process Design: Towards Service-Based Green Information Systems. IFIP 20th World Computer Congress Proceedings, 195-203, Milan, Italy, September 2008.
36. M. Tanelli, *D. Ardagna*, M. Lovera. LPV model identification for Power Management of Web service Systems. Proceedings of the 2008 IEEE Multi-conference on Systems and Control, 1171-1176, San Antonio, Texas, USA, September 2008.
37. *D. Ardagna*, S. Comai, A. Corallo, G.Giunta, L. Iandoli, N. Ingraffia. A Distributed Information System to Coordinate and Support Business Process Execution in a Cluster of Firms. GITMA07 Proceedings (8th Annual Global Information Technology Management Association World Conference), Naples, Italy, June 2007.
38. *D. Ardagna*, G. Giunta, N. Ingraffia, R. Mirandola, B. Pernici. QoS-driven Web Services Selection in Autonomic Grid Environments. GADA06 Proceedings (Int. Symposium on Grid computing, High-Performance and Distributed Applications), LNCS 4276, 1273-1289, Montpellier, November 2006.
39. *D. Ardagna*, C. Cappiello, C. Francalanci, A. Groppi. Brokering Multisource Data with Quality Constraints. ODBASE06 Proceedings (5th International Conference on Ontologies, DataBases, and Applications of Semantics), 807-817, Montpellier, November 2006.
40. J. Almeida, V. Almeida, *D. Ardagna*, C. Francalanci, M. Trubian. Resource Management in the Autonomic Service-Oriented Architecture. ICAC 2006 Proceedings (3rd International Conference on Autonomic Computing), 84-92, Dublin June 2006. **Acceptance rate 21%**. (GGS A++).
41. *D. Ardagna*, C. Cappiello, M. Comuzzi, C. Francalanci, B. Pernici. A Broker For Selecting And Provisioning High Quality Syndicated Data. ICIQ 2005 Proceedings (10th International Conference on Information Quality), 262-279. Boston, November 2005.
42. *D. Ardagna*, M. Trubian, L. Zhang. SLA Based Profit Optimization in Multi-tier Systems. IEEE NCA 2005 Proceedings (4th International Symposium on Network Computing and Applications), 263-266. Boston, July 2005.
43. *D. Ardagna*, C. Francalanci, G. Bazzigaluppi, M. Gatti, F. Silveri, M.Trubian. A Cost-oriented tool to support server consolidation. ICEIS 2005 Proceedings (7th International Conference on Enterprise Information Systems), 323-330. Miami, Florida, May 2005.
44. L. Zhang, *D. Ardagna*. SLA Based Profit Optimization in Autonomic Computing Systems. ICSOC 2004 Proceedings ACM press (2nd International Conference on Service Oriented Computing), 173-182. New York, November 2004.
45. *D. Ardagna*, C. Francalanci, V. Piuri, F. Scotti. Evolutionary Design of Information Systems Architectures. ICAISC 2004 Proceedings (7th International Conference on Artificial Intelligence and Soft Computing), 1-8. Invited paper. Zakopane, Poland, June 2004.
46. *D. Ardagna*, C. Francalanci, M. Trubian. A Cost-oriented Approach for Inftrastructural Design. SAC 2004 Proceedings (19th ACM Symposium on Applied Computing), 1431-1437. Nicosia, Cyprus, March 2004.
47. *D. Ardagna*, C. Francalanci, M. Trubian. A Multi-Model Algorithm for the cost-oriented design of the Information Technology Infrastructure. ECIS 2003 Proceedings (11th European Conference on Information Systems). Naples, June 2003.
48. *D. Ardagna*, C. Francalanci. A Cost-Oriented methodology for the design of Web based IT Architectures. SAC2002 Proceedings (17th ACM Symposium on Applied Computing), 1127-1133. Madrid, March 2002.

Editorial Work

1. D. Ardagna. Performance 2020 Foreword from the General Chair. ACM Performance Evaluation Review. 48(3), 2. December 2020.
2. *D. Ardagna*, C. Cappiello. Springer Computing. Special Issue on Sustainable Computing Systems and Applications. Springer 2017.

3. *D. Ardagna*, A. Brunnert, G. Casale, A. van Hoorn. Proceedings of the second International Workshop on Quality-Aware DevOps. ACM 2016.
4. *D. Ardagna*, A. Brunnert, G. Casale, A. van Hoorn. Proceedings of the first International Workshop on Quality-Aware DevOps. ACM 2015.
5. *D. Ardagna*, M. S. Squillante. SIGMETRICS Performance Evaluation Review Special Issue on Performance and Resource Management in Big Data Applications. ACM DL 2015.
6. *D. Ardagna*, C. Cappiello. Proceedings of Fourth International Conference on Green IT Solutions. SCITEPRESS 2015.
7. *D. Ardagna*, L. Schubert. Springer Computing. Special Issue on Cloud Computing. Springer 2015.
8. *D. Ardagna*, L. Schubert. Proceedings of the 2013 international workshop on Multi-cloud applications and federated clouds. ACM DL 2013.
9. *D. Ardagna*, L. Zhang. Run-time Models for Self-managing Systems and Applications. Springer Autonomic Computing Series 2010.
10. *D. Ardagna*, M. Mecella, J. Yang. Business Process Management Workshops, BPM 2008 International Workshops, Milano, Italy, September 1-4, 2008. Revised Papers. Lecture Notes in Business Information Processing. Springer 2009.

International Workshops

1. M. Precuzzi, F. Filippini, D. Ardagna. Scheduling Deep Learning Jobs Training in the Cloud: Comparing Multiple Approaches. SPARK 2022 Workshop Proceedings. (15th International Workshop on Scheduling and Planning Applications woRKshop). To appear.
2. F. Filippini, D. Ardagna, M. Lattuada, E. Amaldi, M. Ciavotta, M. Riedl, K. Materka, P. Skrzypek, F. Magugliani, M. Cicala. ANDREAS: Artificial intelligence traiNing scheDuler foR accElerAted resource clusterS. EMSICC 2021 Workshop Proceedings. (7th International Workshop on Energy Management for Sustainable Internet-of-Things and Cloud Computing). 388-393.
3. H. Sedghani, *D. Ardagna*, M. Matteucci, G. A. Fontana, G. Verticale, F. Amarilli, R. Badia, D. Lezzi, I. Blanquer, A. Martin, K. Wawruch. Advancing Design and Runtime Management of AI Applications with AI-SPRINT, Position Paper. AIM 2021 Workshop Proceedings (Advances in Artificial Intelligence & Machine Learning: Applications, Challenges & Concerns). 1455-1462.
4. E. Lomurno, S. Samele, M. Matteucci, *D. Ardagna*. Pareto-Optimal Progressive Neural Architecture Search. NEvo@Work ACM Workshop on NeuroEvolution@Work 2021. GECCO (The Genetic and Evolutionary Computation Conference) Companion, 1726-1734. 2021.
5. S. Kalwar, E. Gianniti, J. Y. Kinouani, Y. Ridene, *D. Ardagna*. Performance Degradation and Cost Impact Evaluation of Privacy Preserving Mechanisms in Big Data Systems. INFQ@Valuetools2017 Workshop Proceedings. 82-96. Venice, Italy. 2017.
6. E. Gianniti, *D. Ardagna*, M. Ciavotta, M. Passacantando. A Game-Theoretic Approach for Runtime Capacity Allocation in MapReduce. WACC 2017 Workshops@CCGRID 2017 Proceedings (International Workshop on Assured Cloud Computing and QoS aware Big Data). 1080-1089. Madrid, Spain.
7. M. Ciavotta, E. Gianniti, *D. Ardagna*. Capacity Allocation for Big Data Applications in the Cloud. QUDOS 2017 Workshops Proceedings (3rd International Workshop on Quality-Aware DevOps. 175-176. L'Aquila, Italy.
8. E. Gianniti, A. M. Rizzi, E. Barbierato, M. Gribaudo. Fluid Petri Nets for the Performance Evaluation of MapReduce Applications. INFQ 2016 Workshop Proceedings. 243-250. Taormina, Italy.
9. E. Ataie, E. Gianniti, D. Ardagna, A. Movaghar. A Combined Analytical Modeling Machine Learning Approach for Performance Prediction of MapReduce Jobs in Hadoop Clusters. MICAS-SYNASC 2016 Workshops Proceedings (5th Workshop on Management of resources and services in Cloud and Sky computing). 431-439. Timisoara, Romania.
10. M. Guerriero, M. Ciavotta, G. Paolo Gibilisco, *D. Ardagna*. A Model-Driven DevOps framework for QoS-aware Cloud applications. MICAS-SYNASC 2015 Workshops Proceedings (4th Workshop on Management of resources and services in Cloud and Sky computing). 345-351.Timisoara, Romania.
11. M. Guerriero, M. Ciavotta, G. P. Gibilisco, *D. Ardagna*. SPACE4Cloud: a DevOps environment for multi-cloud applications. QUDOS@SIGSOFT FSE 2015 Workshops Proceedings (4th Workshop on Management of

resources and services in Cloud and Sky computing). 29-30. Bergamo, Italy.
12. G. Casale, *D. Ardagna*, M. Artac, F. Barbier, E. Di Nitto, A. Henry, G. Iuhasz, C. Joubert, J. Merseguer, V. I. Munteanu, J. F. Pérez, D. Petcu, M. Rossi, C. Sheridan, I. Spais, D. Vladušič. DICE: Quality-Driven Development of Data-Intensive Cloud Applications. MiSE 2015 Workshops Proceedings (4th Workshop on Management of resources and services in Cloud and Sky computing). 78-83. Florence, Italy.
13. *D. Ardagna*, M. Ciavotta, R. Lancellotti. A Receding Horizon Approach for the Runtime Management of IaaS Cloud Systems. MICAS-SYNASC 2014 Workshops Proceedings (3rd Workshop on Management of resources and services in Cloud and Sky computing). 445-452. Timisoara, Romania.
14. M. Malekimajd, A. M. Rizzi, *D. Ardagna*, M. Ciavotta, M. Passacantando, A. Movaghar. Optimal Capacity Allocation for executing Map Reduce Jobs in Cloud Systems. MICAS-SYNASC 2014 Workshops Proceedings (3rd Workshop on Management of resources and services in Cloud and Sky computing). 385-392. Timisoara, Romania.
15. G. Rumi, C. Colella, *D. Ardagna*. Optimization techniques within the Hadoop ecosystem: A Survey. MICAS-SYNASC 2014 Workshops Proceedings (3rd Workshop on Management of resources and services in Cloud and Sky computing). 437-444. Timisoara, Romania.
16. M. A. Almeida da Silva, *D. Ardagna*, N. Ferry, J. F. Pérez. Model-Driven Design of Cloud Applications with Quality-of-Service Guarantees: The MODAClouds Approach, MICAS Tutorial. MICAS-SYNASC 2014 Workshops Proceedings (3rd Workshop on Management of resources and services in Cloud and Sky computing). 3-10. Timisoara, Romania.
17. D. Petcu, E. Di Nitto, *D. Ardagna*, A. Solberg, G. Casale. Towards Multi-Clouds Engineering (invited paper). IEEE INFOCOM CrossClouds 2014 Workshop Proceedings. 1-6,Toronto, Canada.
18. E. Di Nitto, M. Almeida, *D. Ardagna*, G. Casale, C. Dorin Craciun, N. Ferry, V. Muntes, A. Solberg. Supporting the Development and Operation of Multi-Cloud Applications: The MODAClouds Approach. MICAS-SYNASC 2013 Workshops Proceedings (2nd Workshop on Management of resources and services in Cloud and Sky computing). 417-423, Timisoara, Romania.
19. M. Miglierina, M. Balduini, N. Shahmandi Hoonejani, E. Di Nitto, *D. Ardagna*. Exploiting Stream Reasoning to Monitor Multi-Cloud Applications. OrdRing 2013 Workshop Proceedings (2nd International Workshop on Ordering and Reasoning). 33-36, Sydney, Australia.
20. M. Miglierina, G. P. Gibilisco, *D. Ardagna*, E. Di Nitto. Model Based Control for Multi-cloud Applications. MiSE 2013 Workshop Proceedings (5th International Workshop on Modeling in Software Engineering). 37-43, San Francisco, USA.
21. D. Franceschelli, *D. Ardagna*, M. Ciavotta, E. Di Nitto. SPACE4CLOUD: A Tool for System PerformAnce and Cost Evaluation of CLOUD Systems. Multi-Cloud 2013 Workshop Proceedings. 27-34. Prague, Czech Republic.
22. S. Benefico, E. Gjeci, R. Gonzalez Gomarasca, E. Lever, S. Lombardo, *D. Ardagna*, E. Di Nitto. The evaluation of CAP Properties on Amazon SimpleDB and Windows Azure Table Storage. MICAS-SYNASC 2012 Workshops Proceedings (1st Workshop on Management of resources and services in Cloud and Sky computing). 430-435. Timisoara, Romania.
23. F. L. Ferraris, D. Franceschelli, M. P. Gioiosa, D. Lucia, *D. Ardagna*, E. Di Nitto, T. Sharif. Evaluating the Auto Scaling Performance of Flexiscale and Amazon EC2 Clouds. MICAS-SYNASC 2012 Workshop Proceedings (1st Workshop on Management of resources and services in Cloud and Sky computing). 423-429. Timisoara, Romania.
24. S. Lombardo, E. Di Nitto, *D. Ardagna*. Issues in handling complex data structures with NoSQL databases. MICAS-SYNASC 2012 Workshop Proceedings (1st Workshop on Management of resources and services in Cloud and Sky computing). 443-448. Timisoara, Romania.
25. *D. Ardagna*, E. Di Nitto, D. Petcu, P. Mohagheghi, S. Mosser, P. Matthews, A. Gericke, C. Ballagny, F. D'Andria, C. Nechifor, C. Sheridan. MODACLOUDS: A Model-Driven Approach for the Design and Execution of Applications on Multiple Clouds. MiSE 2012 Workshop Proceedings (4th International Workshop on Modeling in Software Engineering). 50-56, Zurich, Switzerland.
26. *D. Ardagna*, R. Mirandola, M. Trubian, L. Zhang. Run-time Resource Management in SOA Virtualized Environments. QUASOSS 2009 Proceedings (1st International Workshop on the Quality of Service-Oriented Software System). 39-46, Amsterdam, The Netherlands.
27. J. Anselmi, *D. Ardagna*, P. Cremonesi. A QoS-based Selection Approach of Autonomic Grid Services. SOCP07 Workshop Proceedings (Workshop on Service-Oriented Computing Performance). 1-7, Monterey Bay, California, June 2007.

28. *D. Ardagna*, S. Lucchini, R. Mirandola, B. Pernici. Web Services Composition in Autonomic Grid Environments. GPWW06 Proceedings (International Workshop on Grid and Peer-to-Peer based Workflows). 370-381, Vienna, September 2006.
29. *D. Ardagna*, M. Comerio, F. De Paoli, S. Grega. A Hybrid Approach to QoS Evaluation. EMMSAD 2005 Workshop Proceedings (10th International Workshop on Exploring Modeling Methods in Systems Analysis and Design). 581-592. Porto, June 2005.
30. M. Adorni, F. Arcelli, *D. Ardagna*, L. Baresi, C. Batini, C. Cappiello, M. Comerio, M. Comuzzi, F. De Paoli, C. Francalanci, S. Grega, P. Losi, A. Maurino, S. Modafferi, B. Pernici, C. Raibulet, F. Tisato. The MAIS approach to web service design. EMMSAD 2005 Workshop Proceedings (10th International Workshop on Exploring Modeling Methods in Systems Analysis and Design). 387-398. Porto, June 2005.
31. *D. Ardagna*, B. Pernici. Global and Local QoS Guarantee in Web Service Selection. BPS 2005 Proceedings (Workshop on Business Processes and Services). 32-46. Nancy (France), September 2005.

Book Chapters

1. *D. Ardagna*, M. Ciavotta, G. P. Gibilisco, R. B. Desantis, G. Casale, J. F. Perez, F. D'Andria, R. Sosa Gonzalez. QoS Assessment and SLA Management. Springer, 2016.
2. N. Calcavecchia, *D. Ardagna*, E. Di Nitto. The Emergence of Load Balancing in Distributed Systems: the SelfLet Approach. Run-time Models for Self-managing Systems and Applications, Springer Autonomic Systems Series, 2010.
3. *D. Ardagna.* Process Optimization. Encyclopedia of Database Systems, Springer, 2008.
4. *D. Ardagna*, A. Avenali, L. Baresi, D. Berardi, D. Bianchini, C. Cappiello, M. Comuzzi, V. De Antonellis, F. De Rosa, D. Desideri, C. Francalanci, C. Leporelli, G. Matteucci, A. Maurino, M. Mecella, M. Melchiori, S. Modafferi, E. Mussi, B. Pernici, P. Plebani, D. Presenza. E-Services - Mobile Information Systems - Infrastructure and Design for flexibility and adaptivity, Springer, 2006.
5. *D. Ardagna*, L. Baresi, C. Batini, M. Brioschi, C. Cappiello, M. Comerio, M. Comuzzi, F. De Paoli, C. Francalanci, S. Grega, A. Maurino, S. Modafferi, N. Simeoni. Development of Services for Mobile Information Systems - Mobile Information Systems - Infrastructure and Design for flexibility and adaptivity, Springer, 2006.
6. M. Adorni, F. Arcelli, *D. Ardagna*, L. Baresi, C. Batini, C. Cappiello, M. Comerio, M. Comuzzi, F. De Paoli, C. Francalanci, S. Grega, P. Losi, A. Maurino, S. Modafferi, B. Pernici, C. Raibulet, F. Tisato. The MAIS approach to web service design. Advanced Topics in Database Research - Volume 5, Idea Group, 2006.

Books In Italian

1. F. Amarilli, *D. Ardagna*, M.G. Fugini, R. Tedesco. Impianti Informatici: Tecnologie e Applicazioni. McGraw-Hill 2007.
2. *D. Ardagna,* M.G. Fugini, B. Pernici, P. Plebani. Sistemi Informativi Basati su Web. Franco Angeli 2006.
3. *D. Ardagna*, L. Arisi, A. Biffi, L. Buccoliero, S. Calciolari, C. Francalanci, P. Ganugi, L. Heller, G. Manzoni, A. Martegani,R. Calmieri, F. Timpano, P. Facchi, P. Vercesi. E-government nella Pubblica Amministrazione Lecchese. Franco Angeli 2004.
4. G. Bracchi, C. Francalanci, A. Bognetti, M. Giorgino, T. Monfroglio, *D. Ardagna*, R. Bracchi, C. Cappiello, E. Gilardoni, R. Padovani. La banca multicanale in Europa. Tecnologia, qualità dei servizi e profittabilità. Edibank 2002.

Other Publications

1. B. Guindani, *D. Ardagna*, A. Guglielmi. Bayesian Optimization with Machine Learning for Big Data Applications in the Cloud. SIS 2022. 51st Scientific Meeting of the Italian Statistical Society. Caserta, Italy. 1-6. June 2022. Pearson. To Appear.
2. M. Ciavotta, A. Jahani, M. Lattuada, *D. Ardagna*, E. Amaldi - GPU Scheduling for Deep Learning – International Conference on Optimization and Decision Science (ODS 2019). September 2019. Extended Abstract.

3. M. Scavuzzo, E. Di Nitto, *D. Ardagna*. Experiences and challenges in building a data intensive system for data migration. ICSE 2018, 93. May 2018.
4. M. Ciavotta, *D. Ardagna*, A. Koziolek. Palladio Optimization Suite: QoS optimization for component-based Cloud applications. Valuetools 2015. Extended Abstract. Berlin, Germany.
5. M. Ciavotta, E. Gianniti, *D. Ardagna*, M. Passacantando. Game-Theoretic Approach to Joint Admission Control and Capacity Allocation for MapReduce. AIRO 2015. Pisa, Settembre 2015. Extended Abstract.
6. B. Addis, *D. Ardagna*, G. Carello, A. Capone. Energy-aware joint management of networks and Cloud infrastructures. AIRO 2015. Pisa, Settembre 2015. Extended Abstract.
7. M. Almeida, *D. Ardagna*, N. Ferry, J.F. Peréz. Model-Driven Design of Cloud Applications with Quality-of-Service Guarantees: the MODAClouds Approach. Timisoara, September 2014. Extended Abstract.
8. M. Ciavotta, *D. Ardagna*, M. Passacantando. Generalized Nash Equilibria for the Service Provisioning Problem in Multi-Cloud Systems. AIRO 2014. Como, Settembre 2014. Extended Abstract.
9. D. Molinari, M. Mazuran, C. Arias, G. Minucci, F. Atun, *D. Ardagna*. PoliRisposta: Overcoming present limits of flood damage data. European Geosciences Union General Assembly 2014. Vienna, April 2014. Extended Abstract.
10. J. Anselmi, *D. Ardagna*, J. C. S. Lui, A. Wierman, Y. Xu, Z. Yang. The economics of the cloud: price competition and congestion. NetEcon 2013. Pittsburgh, June 2013. Extended Abstract.
11. *D. Ardagna*, M. Caldirola. Tecnologie di virtualizzazione: evoluzione in atto. AEIT, 16-21, Luglio-Agosto 2010. Invited Paper.
12. *D. Ardagna*, L. Baresi, S. Comai, M. Comuzzi, B. Pernici, M. Pianciamore, S. Prandin. Una Piattaforma di Servizi Avanzata a Supporto di Distretti Cooperativi: il progetto DISCoRSO. ITAIS 2007. Venezia, Ottobre 2007.
13. *D. Ardagna*, M. Giorgetta, M. Trubian. Optimum Servers Partitioning in Autonomic Computing Infrastructure. AIRO 2006. Cesena, Settembre 2006. Extended Abstract.
14. *D. Ardagna*, M. Trubian. Resource Allocation Optimization in Multi-tier Data Centers with Strict QoS Performance Guarantees. AIRO 2005. Camerino, Settembre 2005. Extended Abstract.
15. *D. Ardagna*, B. Pernici. Global and Local QoS Constraints Guarantee in Web Service Selection. ICWS 2005 Poster Track Proceedings (3rd IEEE International Conference on Web Services), 462-463. Orlando, July 2005.
16. L. Zhang, *D. Ardagna*. SLA Based Profit Optimization in Web Systems. WWW 2004 Poster Track (13th International World Wide Web Conference), 462-463. New York, May 2004.
17. *D. Ardagna*, C. Francalanci, E. Zavarise. Internet sostituirà le attuali infrastrutture telefonica e televisiva? SMAU Ricerche. March 2002. Invited Paper.
18. *D. Ardagna*, L. Bazzardi, P. Facchi, C. Francalanci. Internet, standardizzazione tecnologica e riduzione dei costi per le imprese: il primo passo verso ritorni positivi. SMAU Ricerche. March 2002. Invited Paper.
19. *D. Ardagna*, C. Francalanci, M. Trubian. Cost oriented optimization of information technology (IT) infrastructures. AIRO 2002. L'Aquila, Settembre 2002. Extended Abstract.

Impact of Selected Publications

1. F. Filippini, M. Lattuada, M. Ciavotta, A. Jahani, *D. Ardagna*, E. Amaldi. A Path Relinking Method for the Joint Online Scheduling and Capacity Allocation of DL Training Workloads in GPU as a Service Systems. IEEE Transactions on Services Computing. To Appear. 1-16. 10.1109/TSC.2022.3188440. (Scimago Q1).
   Contribution: main publication of my PhD student's work. Led the problem definition and research agenda. Participated to the optimization problem formulation and validation.
   Impact: Reference paper of the TETRAMAX ANDREAS technology transfer project and TETRAMAX award, proposes a joint scheduler and resource allocator for deep learning jobs. Provides 95% costs savings in comparison to an IBM Research prototype.

2. M. Ciavotta, G. P. Gibilisco, *D. Ardagna*, E. Di Nitto, M. Lattuada, M. A. Almeida da Silva. Architectural Design of Cloud Applications: a Performance-aware Cost Minimization Approach. IEEE Transactions on Cloud Computing. To Appear. 1-18. DOI: 10.1109/TCC.2020.3015703. (Scimago Q1).

Contribution: main publication of my PhD student's work. Led the research agenda, the optimization problem formulation and validation.
Impact: Applied to an industry use case, provides between 21% and 85% costs savings in comparison to a SOTA alternative method.

3. M. Lattuada, E. Barbierato, E. Gianniti, *D. Ardagna*. Optimal Resource Allocation of Cloud-Based Spark Applications. IEEE Transactions on Cloud Computing. 10(2), 1301-1316. 2022. (Scimago Q1).
Contribution: led the research agenda, the optimization problem formulation, the design of the heuristic method and validation.
Impact: global optimum solution gap reduced by 30% with respect to a SOTA method by Spark inventors.

4. E. Gianniti, M. Ciavotta, *D. Ardagna*. Optimizing Quality-Aware Big Data Applications in the Cloud. IEEE Transactions on Cloud Computing. 9(2), 737-752. 2021. (Scimago Q1).
Contribution: main publication of my PhD student's work. Led the research agenda. Worked on the performance models and participated to the optimization problem formulation.
Impact: global optimum solution gap reduced by 10-30% with respect to a SOTA method by Spark inventors.

5. *D. Ardagna*, M. Ciavotta, R. Lancellotti, M. Guerriero. A Hierarchical Receding Horizon Algorithm for QoS-driven control of Multi-IaaS Applications. IEEE Transactions on Cloud Computing. 9(2), 418 - 434. 2021. (Scimago Q1).
Contribution: led the research agenda and the validation activities (cloud experiments and simulation). Participated to the optimization problems receding-horizon formulation.
Impact: one of the few literature works providing a multi-time scale solution for resource management. With respect to heuristic provided by previous research works cost-savings range in 50-100%.

6. A. Maros, F. Murai, A. P. Couto da Silva, J. M. Almeida, M. Lattuada, E. Gianniti, M. Hosseini, *D. Ardagna*. Machine Learning for Performance Prediction of Spark Cloud Applications. CLOUD 2019: 99-106. Acceptance rate 20.8%. (GGS A-).
Contribution: responsible for the modelling of the computing/cloud aspects of the paper. Co-led the research agenda with Prof. Almeida an led the validation activities.
Impact: Significantly improved the accuracy of Spark inventors' performance models reducing the percentage error from 126-187% to only 5-19% especially under resource contention and performance variability.

7. *D. Ardagna*, M. Ciavotta, M. Passacantando. Generalized Nash Equilibria for the Service Provisioning Problem in Multi-Cloud Systems. IEEE Transactions on Services Computing. 10(3), 381-395. 2017. (Scimago Q1).
Contribution: led the research agenda and participated to the game formulation. Led validation activities setup.
Impact: Extension of my work in [selected publication 9] to multi-cloud scenarios (79 GoogleScholar citations).

8. B. Addis, *D. Ardagna*, B. Panicucci, M. Squillante, L. Zhang. A Hierarchical Approach for the Resource Management of Very Large Cloud Platforms. IEEE Transactions on Dependable and Secure Computing. 10(5), 253-272, 2013. (Scimago Q1).
Contribution: summary of my ten years of work on Green IT. Led the research agenda and the definition of the hierarchical multi-time scale optimization problem. Led validation activities setup.
Impact: one of the very few literature works (104 GoogleScholar citations) implementing resource management solutions which provides both performance and availability guarantees.

9. *D. Ardagna*, B. Panicucci, M. Passacantando. Generalized Nash Equilibria for the Service Provisioning Problem in Cloud Systems. IEEE Transactions on Services Computing. 6(4), 429-442, 2013. (Scimago Q1).
   Contribution: led research agenda. Participated to game model formulation. Developed the toolchain used for the validation.
   Impact: reformulation of my work in [selected publication 11] as a potential game, providing a fully distributed implementation. Compared to other SOTA solutions Price of Anarchy improved by 50-70% (139 GoogleScholar citations).

10. *D. Ardagna*, B. Panicucci, M. Trubian, L. Zhang. Energy-Aware Autonomic Resource Allocation in Multi-tier Virtualized Environments. IEEE Transactions on Services Computing. 5(1), 2-19, 2012. (Scimago Q1).
    Contribution: led research agenda. Participated to the optimization model formulation. Coordination of the development of the main tool used for the validation. Performed validation activities.
    Impact: one of the very first works (259 GoogleScholar citations) on energy-aware resource management of cloud data centers. Improvement on average by 45% of the solution provided by the IBM Tivoli middleware.

11. *D. Ardagna*, B. Panicucci, M. Passacantando A Game Theoretic Formulation of the Service Provisioning Problem in Cloud Systems. WWW 2011 Proceedings (20th International World Wide Web Conference). 177-186. Hyderabad, India. Acceptance rate 12.3%. (GGS A++).
    Contribution: led research agenda, developed the main tool used for the validation. Participated to the game model formulation.
    Impact: one of the very first works (179 GoogleScholar citations) which applied Game Theory to the management of cloud services identifying Generalized Nash Equilibria.

12. *D. Ardagna*, B. Pernici. Adaptive Service Composition in Flexible Processes. IEEE Transactions on Software Engineering. 33(6), 369-384, 2007 (top ranked paper, 1063 citations). (Scimago Q1).
    Contribution: main output of my post-doc research. Worked on the optimization problem formulation and negotiation framework setting. Toolchain development and execution of the experimental campaign.
    Impact: one of the most cited literature works (1205 GoogleScholar citations) for solving Web Services composition problem with global quality of service constraints.

Research Statement

On-going Research Directions and Recent Achievements

My research work focuses on the design, prototype, and evaluation of **resource management algorithms** for **large scale distributed systems** in several application areas ranging from web applications to **big data** and **Artificial Intelligence** (AI) in **fog/edge/cloud** infrastructures.

In particular, my work aims at: i) **modelling** applications and ICT systems **performance**, and ii) designing **optimization algorithms** for the **management** of **computing continua resources**.

Nowadays, applications are built of top of **very complex software stacks** and are often characterized by heterogeneous and irregular data accesses and computation patterns on the top of massively parallel algorithms. Given their pervasiveness, ICT systems need to provide performance guarantees to their end-users (in some areas, real-time or near-real-time) and **solutions to make intelligent resource management decisions** (e.g., when to move computation from the edge to the cloud, considering the network capabilities as well as the security and/or sensitivity of data) need to be developed.

The **performance prediction** of a given application in a target configuration becomes then a **key task** to support the **proper planning** (at design time) and **management** (at runtime) of the available **resources**.

Recently, I focused on the performance modelling of **big data systems** and **AI applications** applying several analytical

techniques ranging from **queuing networks** to **stochastic Petri Nets**[1,2] to supervised **machine learning** (ML)[3].

ML is a hot topic today and it demonstrated to be very effective since, given its black-box approach and capability to learn systems behavior from data, it allows to **predict applications performance without** the knowledge of the **internals** of the underlying systems. Specifically, I evaluated the accuracy which can be achieved through **several regression techniques** (ranging from linear regression to XGBoost) of Hadoop, Spark and AI applications investigating the effectiveness of the different features that can be extracted from the application (e.g., size of the input data/training set, batch size, number of tasks in a stage, etc.) and from the platform configuration (e.g., number of cores/Virtual Machines/GPUs, I/O buffers size, etc.).

I generally obtained very accurate models but the most significant results I achieved are related to the study of Spark performance, one of the most popular big data processing engines used by the industry today. My approach demonstrated to be very effective, significantly improving the accuracy of the performance models of Spark inventors. Indeed, Spark inventors' work can accurately estimate the performance of very regular applications, but its accuracy falls short when **applications** present more **irregular behavior** (due to the application characteristics itself, e.g., alternating I/O bound and CPU bound stages, or due to resource contention) or extrapolation capabilities are tested (e.g., training the performance regression model on small data sets or few cores, predict the performance for large data sets/many cores). In contrast, the approach I developed demonstrated to be able to address more **complex scenarios** by providing highly accurate predictions (**largest error less than 20%**[3]).

This research work led to also to the development of an **open-source** ML library (**aMLLib**) initially developed within the **H2020 ATMOSPHERE project** which is currently used within the **H2020 AI-SPRINT** and **H2020 EuroHPC LIGATE** projects I am involved-in.

More **accurate performance models** also allowed to develop more **effective capacity planning** and **resource management** solutions. The most important results I achieved in this research area lay in **design time solutions** for **Hadoop** and **Spark** applications and **runtime management** of **Spark** applications.

In the first scenario, I led the development of **D-SPACE4Cloud**, a novel open-source tool for the **right-sizing of cloud clusters**, integrated in a DevOps-inspired approach developed within the **H2020 DICE** project. D-SPACE4Cloud explores the space of alternative cloud configurations, seeking the minimum cost deployment that satisfies quality of service constraints. The capacity planning problem has been formulated by means of a mathematical model, with the aim of minimizing the cost of cloud resources. The (NP-hard) problem considers several Virtual Machine (VM) types as candidates to support the execution of Hadoop/Spark applications from multiple users sharing the same cloud cluster. We developed a **parallel** and **efficient simulation-optimization procedure** able to determine an optimized configuration. Hadoop and Spark **performance** are estimated by relying on **multiple models**, including simulation based on queueing networks, stochastic Petri nets, an ad-hoc discrete event simulator (dagSim designed by **Prof. Marco Gribaudo's group** within the **H2020 EUBRA-BIGSEA** project I led for Politecnico di Milano research unit) and ML. Results have shown that the **global optimum solution gap** can be **reduced** by **10-30%** with respect to an **exhaustive method** that relies of **Spark inventors'** performance model.

In the second scenario, I led the development of a set of **run-time optimization-based resource management policies** for Spark applications running in cloud clusters characterized by hard and soft deadlines (**OPT_IC**, **OPT_JR** open-source tools). The optimization policies address two scenarios: i) identification of the **minimum capacity** to run a Spark **application within the deadline**; ii) **re-balance** of the cloud **resources** in case of **heavy load**, minimizing the soft deadline application tardiness. The solution we developed relies on an initial non-linear programming model formulation and a very fast local search exploiting - at the same time - a gamut of performance modelling techniques, trading-off model accuracy and execution time. The results we obtained demonstrate that the **percentage error** of the prediction of the **optimal resource usage** with respect to **system measurement** and **exhaustive search** is the range **4-29%** while an **exhaustive search** based on **Spark inventors**' performance model presents an average error in the range **6-63%**. The proposed approach demonstrated to be able to address complex problems like computing the optimal redistribution of resources among **tens of applications** in a **prototype cluster**, in **less than a minute** with an error of **8%** on average. On the same considered tests, **Spark inventors'** approach obtains an average error of about **57%**.

Recently, I am also focusing on **AI applications** and, in particular, I modelled the **performance** of **training** and **inference tasks** of **Deep Learning** (DL) models (relying also in this case on **ML techniques**) and I developed **advanced scheduling**

[1] M. Lattuada, E. Barbierato, E. Gianniti, *D. Ardagna*. Optimal Resource Allocation of Cloud-Based Spark Applications. IEEE Transactions on Cloud Computing. 10(2), 1301-1316. 2022.
[2] E. Gianniti, M. Ciavotta, *D. Ardagna*. Optimizing Quality-Aware Big Data Applications in the Cloud. IEEE Transactions on Cloud Computing. 9(2), 737-752. 2021.
[3] A. Maros, F. Murai, A. P. Couto da Silva, J. M. Almeida, M. Lattuada, E. Gianniti, M. Hosseini, *D. Ardagna*. Machine Learning for Performance Prediction of Spark Cloud Applications. CLOUD 2019: 99-106. Acceptance rate 20.8%.

solutions to support **DL training jobs** running on **GPU based clusters**[4].

DL algorithms are used to tackle complex problems that require to process massive datasets to train Neural Networks before they achieve reasonable prediction accuracy and generality. The introduction of the general purpose computation on GPUs, providing an interface to massive parallelism, has significantly extended the set of previously intractable problems that can be solved within a reasonable computing time. Consequently, the market growth for GPU services is expected to be quite remarkable in the next years, starting from over 700 million USD in 2018 and increasing with a compound annual rate of over 38% up to 2025[5]. GPU acceleration is particularly suited to DL training tasks, providing a performance boost from 5 to 40x with respect to CPU-based systems[6].

Despite those great benefits, GPU-based servers are characterized by considerably high costs (about 200k USD for high-end systems like NVIDIA DGX A100) while GPU-based VMs are still remarkably expensive, being characterized by a price 5-8x higher than those of VMs exploiting only CPUs. As a consequence, the problem of determining an efficient schedule for DL training jobs and other applications that are to be deployed on GPU-based systems has great importance. In my research I developed (in **collaboration** with the DEIB **Operations Research group**) an advanced scheduler (at TRL 8 funded by the **H2020 TETRAMAX ANDREAS** project) to support the execution of multiple continuously submitted DL training jobs in **private** or **public cloud clusters**[4]. Individual nodes can be configured from various VM types each one featuring, possibly, several and heterogeneous GPUs. More than one job can run on the same node, and, in this case, the available resources are partitioned and statically allotted to avoid interference. Jobs are characterized by a due date, a tardiness cost (i.e., a penalty cost proportional to the difference between the job completion time and the due date), and a priority. Finally, job preemption is also allowed to manage higher priority submissions. The resulting **online resource allocation and scheduling problems** aim at minimizing the overall job execution costs (the sum of the costs incurred to run the VMs in public clouds or **energy costs** in **private clouds**, plus the tardiness penalty costs). Online decisions concern the VM type selection for each computing node, the order in which the jobs are executed, and how resources are partitioned and assigned to each job. The problem has been modeled through mixed integer linear programming and it is tackled by developing a heuristic approach based on randomized greedy and path relinking. This method is specifically designed and implemented to be efficient while achieving high-quality solutions. Our experimental campaign demonstrated the effectiveness of our method in practical scenarios, since systems with up to **100 nodes, 800 GPUs and 450 concurrent jobs** can be optimized in less than **7 seconds**. Significant costs savings are attained with respect to **first-principle methods** (based on Earliest Deadline First, First In First Out, Priority Scheduling), with an average percentage gain **between 23 and 97%** achieving from **20 to 80%** cost reduction with respect to a solution developed by **IBM Research**. The validation of the proposed approach in a cloud prototype environment based on **Microsoft Azure** and on an **ARM based HPC system** showed a deviation below **5%** and **12%**, respectively, between real and predicted costs[7].

Short- to mid-term Research Directions (up to 5 years)

In the short to mid-term, I will still be working on **performance modelling** and **resource management** of **AI** applications but taking a broader perspective with respect to what I achieved so far. AI software platforms European market forecast to grow significantly through 2024, approaching USD 4 billion in revenue[8]. Many of the benefits of this evolution will come from **using computing resources** at the **periphery of the network**, i.e., where source data is produced. Many companies are evaluating the use of edge computing for data collection, processing, and online analytics to reduce applications latency and data transfers. A growing number of use cases, e.g., predictive maintenance, machine vision, and healthcare to name a few, can benefit from AI applications spanning edge-to-cloud infrastructures leveraging resources available at the **computing continuum**. **Edge intelligence**, i.e., edge-based inferencing, will become the **foundation** of **all industrial AI applications** while most new applications will involve some AI components at various levels of the continuum. Training and/or retraining AI models at the edge will provide opportunities to optimize the use of computational resources, preserving data privacy and increasing data security. In this context, the EU Commission has set a Coordinated Plan to boost investments through the next

[4] F. Filippini, M. Lattuada, M. Ciavotta, A. Jahani, *D. Ardagna*, E. Amaldi. A Path Relinking Method for the Joint Online Scheduling and Capacity Allocation of DL Training Workloads in GPU as a Service Systems. IEEE Transactions on Services Computing. To Appear. 1-16. DOI: 10.1109/TSC.2022.3188440.

[5] GlobalMarketInsights. Gpu as a service market size by product. https://www.gminsights.com/industry-analysis/gpu-as-a-service-market

[6] S. Madougou et al. The landscape of GPGPU performance modeling tools. Parallel Computing. 56, 18–33. 2016.

[7] F. Filippini, *D. Ardagna*, M. Lattuada, E. Amaldi, M. Ciavotta, M. Riedl, K. Materka, P. Skrzypek, F. Magugliani, M. Cicala. ANDREAS: Artificial intelligence traiNing scheDuler foR accElerAted resource clusterS. EMSICC 2021 Workshop Proceedings. (7th International Workshop on Energy Management for Sustainable Internet-of-Things and Cloud Computing). 1-6. To Appear.

[8] IDC Edge Spending Guide, V2 2020 (January 2021).

funding programme Digital Europe (2021-2028) to help drive European AI leadership.

Even if today the huge interest on AI by Academia and Industry is due to its expected impact on society and business opportunity, in production environments (e.g., at Google, Facebook, Uber, and other big players' premises), AI/ML code is only 10% of the system, which also includes the code to deploy, configure, monitor, manage the infrastructure and trigger models retraining[9]. In my research, I will work to define a **novel framework** for **developing** and **operating AI applications** together with their data, exploiting computing continuum environments. This is the main goal of the **H2020 AI-SPRINT** (Artificial intelligence in Secure PRIvacy-preserving computing continuum) recently funded project (H2020-ICT-2020-40, *Cloud Computing: towards a smart cloud computing continuum*) started January 1, 2021 which will offer novel tools for AI applications development, secure execution, easy deployment, as well as runtime management and optimization. AI-SPRINT tools will allow **trading-off application performance** (in terms of end-to-end latency or throughput), **energy efficiency** (one of the main **Horizon Future** objective), and **AI models accuracy** while providing **security** and **privacy guarantees**. The AI-SPRINT framework will support AI applications data protection, architecture enhancement, agile delivery, runtime optimization, and continuous adaptation. I am the **Project Coordinator** of the AI-SPRINT project which also involves **Prof. Matteo Matteucci's AI group** (leading the use cases), **Prof. Giacomo Verticale's group** (responsible for 5G network runtime management and security), **Cefriel** (involved for integration activities) and **Fondazione Politecnico di Milano** (supporting in management tasks).

From a research standpoint, in AI-SPRINT I will be directly involved in: i) to develop **novel performance models** for AI DL models inference (possibly running across multiple resources along the computing continuum and by relying on remote GPUs in disaggregated hardware environments), ii) to design novel tools supporting **AI applications design** (in particular to identify the optimal component placement and computing continuum resource selection), iii) to identify **optimal DL model partitioning** (i.e., run a deep network model along the computing continuum taking into account privacy issues and multiple deployment options according to edge devices and network capabilities), iv) to develop **novel network architecture search algorithms** (to identify the Pareto front among multiple design objectives, i.e., AI model accuracy, application end-to-end performance, energy, etc.), v) to design **novel algorithms** for the runtime management of the computing continuum resources (to cope with workload fluctuations and to avoid edge resources saturation).

I will tackle the AI inference **performance modelling** with ML techniques also developing tools to **automate the profiling activities** across the **computing continuum** and automating the **model selection** and **hyperparameters tuning** by extending the aMLlib library beyond grid-search relying on novel **Bayesian optimization methods**. For **AI applications design**, I plan to develop solutions for AI applications components placement supporting the seamlessly and secure execution of components across distributed heterogeneous infrastructures, including virtualized, container-based resources, GPGPUs, multi-clouds (possibly based on the Function as a Service paradigm), and AI-enabled sensors. I will consider, through mixed integer non-linear programming modelling and heuristic methods, the **joint problem of component placement**, **DL model partition selection**, and **resource selection** (to address questions like: is it more convenient to adopt a drone with a GPU board for a predictive maintenance application or offload the DL model inference task to the tablet of the drone pilot?). **Optimal model partitioning** will face the problem to identify - at design time - multiple candidate deployments for a DL model, considering multiple network configurations connecting edge and cloud, memory and computing constraints of AI-enabled sensors and the privacy constraints of the injected data.

Neural Architecture Search (NAS) is the process of automating architecture engineering, searching for the best deep learning configuration. This is an important research problem with a significant industry impact, since DL models are usually handcrafted and obtained through a very time-consuming process (AutoML, the Google NAS-cloud based solution is one of the most expensive Google service, priced at 18$/hour). Current literature solutions focus mainly on the target model accuracy (and the core problem is to predict the accuracy through a short training in a subset of the training set) neglecting inference execution time, (re)training time of the model, the resource constraints, and the possibility to partition the model across the continuum resources. In my research, I will extend state of the art NAS algorithms integrating performance predictors to the existing approaches to carry out a joint prediction of time and accuracy for each candidate neural network, searching through a Pareto front. **Performance modelling, model partitioning, and NAS** research activities will be tightly integrated facing the problems with a systemic approach. Finally, for the **runtime management** of the **computing continuum** resources, extremely challenging given the strict re-configuration time required to avoid application downtime and the edge limited resources, I will develop **fast heuristics methods** tailored with the AI applications characteristics and integrated within the AI-SPRINT

[9]D. Sculley et al. Hidden Technical Debt in Machine Learning Systems, Google. NIPS 2015.

framework (e.g., exploiting the design time solution, re-evaluating at runtime a change in the partition point of a DL model). As final task in AI-SPRINT, I plan also to implement **resource management policies by relying on AI**, and in particular, on **Reinforcement Learning** (RL) methods exploiting the RL characteristics to adapt to application or end-users' behavior evolution and integrating my expertise in system modelling to speed-up RL learning phase developing **model-based learning solutions**.

The last objective I plan to include in my research agenda is to **focus** also on the **evolution of AI applications** and their integration with **High Performance Computing** (HPC) systems and big data research focusing on **performance** and **energy optimization**. The **European HPC program** is fostering the **convergence** of **AI**, High Performance Data Analytics (**HPDA**) and **HPC**. AI and HPC are converging on one hand to support the training of DL models of exponentially increasing size (AI models double their computing requirements every 3.5 months[10]) and, on the other hand, AI is integrated in traditional HPC workflows to approximate the results of computing intensive scientific tasks through inference. AI and big data analytics are also intertwined, sharing the need for very large data set, data cleaning and curation and compute intensive training. Note that, at this scale, AI model training has an important energy footprint. For example, training of the GPT-3 model with its 175 billion parameters[11] required 190 MWh corresponding 85 tons of CO2 equivalents, the same amount produced by a new European car driving 700,000 km (i.e., driving back and forth to Moon).

Nowadays, most of the **software running on HPC clusters has complex structure** and depends on **many parameters**. The ability to optimally choose values for such parameters holds utmost importance. Indeed, cluster configuration can impact the performance significantly. Sensible parameter values can allow obtaining solutions to complex problems in a reasonable amount of time, as well as **decreasing economic cost** due to the reduced computational load and **energy impact**. In particular, I plan to apply **Bayesian Optimization** (BO) methods (**in collaboration with Prof. Alessandra Guglielmi** full Professor in Bayesian Statistics at DMAT) to optimize the configuration of AI models running in HPC clusters. The adoption of BO is promising since the running time of an HPC program/AI training job can be modeled as a black-box objective function to be minimized over the feature space. BO does not require many samples to perform predictions and, given its black-box nature, it ignores the internal structure of the system to perform inference. This is suitable both due to the complex structure of HPC-running software and due to the need of small sample sizes, which greatly limits the amount of knowledge one can gain on the objective function. In particular, I plan to extend the ongoing work on the performance auto-tuning of drug discovery applications running in HPC clusters within the **H2020 EuroHPC LIGATE** project (**led** for PoliMI research unit by **Prof. Gianluca Palermo** and **Prof. Cristina Silvano**) to optimize the hyper-parameter tuning and training of very large AI models. The final goal is to make AI sustainable in line with the objectives of the *Horizon Europe pillar 2 Global Challenges and industrial competitiveness, cluster 4 - Digital, industry and space.*

Teaching Statement

Short- to mid-term Teaching Plan (up to 5 years)

I am currently teaching two significantly different courses at the Master level for 15 credits within the Mathematical Engineering and Computer Science and Engineering programs. In the short to mid-term, I plan to evolve slightly the two master courses and to propose a new PhD course on my research topics.

I have been teaching **Algorithms and Parallel Computing** in the **Mathematical Engineering** study programme since 2016. The course had some historical structural problems due to a huge gap between the students' starting knowledge and the learning objectives setup by the Faculty of the Master study programme. Indeed, students have a basic computer science and programming background (10 credits taught within Informatica A attended at the first year at the Bachelor level) while the goals of the course are to introduce the C++ language, to provide an introduction to parallel programming through MPI and (in the 2016 to 2019 editions) to provide also a basic introduction to Python and Spark. I have been continuously working in reshaping the course material (started almost from scratch in 2016, writing also an exercise book) and given the reduction of the number of teaching hours (from 120 of the first year to 100), I had to drop the most advanced topics (template programming in C++ and big data technologies). The course is now in a good shape and was well received by students during the last edition (students' evaluation 3.1/4).

The Faculty of the Master study programme is aware of the course criticalities and, given the need of Mathematical

[10] Stanford University 2019 AI Index Report.

[11] T. B. Brown et al. Language Models are Few-Shot Learners. NeurIPS 2020.

Engineering students to strengthen their computer science skills (especially needed for the statistic/data science and finance specializations), from the next Academic year additional 10 credits on computer science topics had been provided at the third year of the Bachelor level last year (Algorithms and Architectures for High Performance Computing, taught by Prof. Fabrizio Ferrandi). This will give me the opportunity to reduce the basic C language notions recall that requires almost the first three weeks of the course and to add new topics on parallel programming. I plan to extend the course by including OpenMP which, on one side given its simplicity with respect to MPI, will allow me to provide a smoother introduction to parallel programming and, on the other side given the possibility of OpenMP to offload computation to GPUs, has an important interest and impact on the research of the Math Department which is oriented to HPC.

I also teach **Computing Infrastructures** within the **Computer Science and Engineering** study programme. I have been already involved in this course (I taught the 2019-20 edition, I taught one edition of the Computer Systems in Como which used to include the same topics, and I was lecturer of Prof. Gribaudo and Prof. Mirandola in 2015-16 edition) which was always well received by students. Fifty percent of this course is devoted to introducing queueing network theory and evaluation of systems availability basics while the remaining part introduces enterprise data centers, virtualization and cloud computing. The first part of the course requires a slightly update of the teaching material (mostly based on the Lazowska's book dated 1984 which is a good book but provides examples that do not match with the current technology). Even if I like a lot the first part of the course since it is very close to my PhD studies and my initial research path, the second part, which is related to IT technology, is the one I like more. This second part has to continuously evolve given also the technology trends and the industry needs. In the short-term, I plan to add containers technologies and Kubernetes, given their wide adoption in the industry today, and push a bit more the emphasis from Infrastructure as a Service Systems to Platform as a Service and introducing Function as a Service. I plan to add these new topics through hands-on and inviting industry speakers.

Finally, I recently proposed a new **PhD course** on **AI applications design and runtime management in computing continua** that will be delivered on February 2023. This course will be based on the experience of the AI-SPRINT project which, as part of its dissemination and exploitation activities, will provide a one-week **MOOC** available through Coursera or a YouTube project channel and a Master course at Cefriel. The course will introduce the basic "standard" technologies (e.g., Pytorch, Keras, etc.) but will also include the advanced frameworks under development that will be available as PoCs at research level. The course will also involve AI-SPRINT faculty and industry partners.

In particular, the application **design** part will be delegated to **Barcelona Supercomputing Center** researchers who are very active in defining novel programming models for HPC and edge/fog systems. **Cloud orchestrations, Function as a Service** will be taught by **University of Valencia** partners who developed TRL 9 solutions on these topics and provide consultancy services in the eastern Spain area and can bring several **industry case studies**. The course will include also a part devoted to **computing continua security** that will involve the **Technical University of Dresden** colleagues and **their start up** that provides a tool to support transparent application deployment and execution on Trusted Execution Environment based systems. **TinyML** principles and **IoT hardware platforms** will be introduced by exploiting our links to **STMicroelectronics** while the part related to **performance modelling**, **DL models partitioning**, **NAS**, **privacy**, and **5G networks** will be also taught by the DEIB colleagues directly involved in the project (**Prof. Matteo Matteucci** and **Prof. Giacomo Verticale**). The PhD course will be taught in **blended mode** mixing: i) **in person** classes by DEIB faculty members and, possibly, **visiting partners** (with a focus on the core topics and tools training), ii) **remote teaching** by AI-SPRINT partners, iii) the **MOOC material** (for the basic introductory concepts).