Gianmarco
De Francisci Morales

Curriculum Vitae (March 5, 2013)

Experience

Vocational

Oct'11–Present **Postdoctoral Researcher**, *Yahoo! Research*, Barcelona, Spain.
Research on Web and social network mining.

Aug '11–Present **Apache Pig Committer**, *Apache Software Foundation*.
Committer for the Apache Pig project from the Hadoop ecosystem. Member of the open source community that develops Pig, a framework for large scale data analysis.

Oct '12–Dec '12 **Visiting Scientist**, *Yahoo! IntoNow*, Sunnyvale, USA.
Design and development of a backend for stream processing based on Storm.

Apr '12–Aug '12 **Google Summer of Code Mentor**, *Google & Apache Software Foundation*.
Mentor for Apache Pig in the GSoC project. Implementation of a distributed RANK function in MapReduce. Integrated in Pig 0.11. code

Nov '09–Jun '12 **Research Associate**, *National Research Council (ISTI-CNR)*, Pisa, Italy.
Research on cloud computing for large scale data analysis.

Oct '10–Sep '11 **Research Intern**, *Yahoo! Research*, Barcelona, Spain.
Research on Data Intensive Scalable Computing.

May '11–Aug '11 **Google Summer of Code Student**, *Google & Apache Software Foundation*.
Student in the GSoC project for Apache Pig. Sugar for Pig: syntactic sugar features to simplify the use of Pig (Java, ANTLR). Integrated in Pig 0.10. code 1 code 2

May '10–Aug '10 **Google Summer of Code Student**, *Google & Apache Software Foundation*.
Student in the GSoC project for Apache Pig. Design and implementation of a binary comparator for secondary sort. The new comparator is 10x faster, and produces a 20% end-to-end improvement for relevant queries (Java). Integrated in Pig 0.9. code

May '07–Feb'09 **System Administrator**, *Consorzio COMETA*, Catania, Italy.
Administration of Linux clusters for High Performance Computing (HPC). Management of grid middleware (gLite). Porting, integration and development of grid-enabled and parallel (MPI) applications.

Apr '04–Jul '04 **Intern**, *National Institute for Nuclear Physics (INFN)*, Catania, Italy.
Design and development of a framework for booking and allocation of resources in a Grid environment (Java).

Miscellaneous

Jun '02–Sep '02 **Personal Teacher**.
One to one teaching of mathematics, algebra and geometry.

✆ (+34) 6345 14199 • ✉ gdfm@gdfm.me • ☝ gdfm.me

Education

Mar '09–Mar'12   **Ph.D. ("Doctor Europaeus") in Computer Science and Engineering**, *IMT Institute for Advanced Studies Lucca*, Italy, Full marks and honors (*Excellent*).
Thesis: "Big Data and the Web: Algorithms for Data Intensive Scalable Computing".

Oct '04–Apr '08   **Master's in Computer Engineering**, *University of Catania*, Italy, Full marks and honors (*110/110 cum laude*).
Thesis: "Multi-objective design space exploration of embedded systems in a Grid environment" (developed in C++, written in English).

Mar '06–Aug '06   **Erasmus student**, *Pazmany Peter Catholic University*, Budapest, Hungary.
Artificial Intelligence. Design Patterns in Java. Cellular Neural Networks.

Oct '01–Oct '04   **Bachelor's in Computer Engineering**, *University of Catania*, Italy, Full marks and honors (*110/110 cum laude*).
Thesis: "Serivzi di allocazione e monitoraggio dei job e delle risorse di una Grid" (Allocation and monitoring services for jobs and resources in a Grid environment, Java).

Courses

Oct '11–Dec '11   **Introduction to Artificial Intelligence**, *Stanford University*, ai-class.com.
First Massive Online Open Course (MOOC) on Artificial Intelligence, by Sebastian Thrun and Peter Norvig.

Oct '11–Dec '11   **Introduction to Machine Learning**, *Stanford Univesrity*, ml-class.com.
First Massive Online Open Course (MOOC) on Machine Learning, by Andrew Ng.

Jul '10   **Lipari Summer School on Computational Complex Systems**, *Jacob T. Schwartz International School for Scientific Research*, Lipari, Italy.
Social Network Mining: Graph mining; Social network analysis; Dynamical processes on complex networks; Rise of cooperation in society with greed.

Jul '09   **CASE Summer School**, *Free University of Bozen*, Bolzano, Italy.
Software Engineering: Agile software process management; Test Driven Development (TDD); Continuous testing of configurable systems; Software and process measurement.

Dec '08   **Matlab Official Course**, *The MathWorks*, Palermo, Italy.
Matlab: Fundamentals; GUI design; Integration with external languages; Distributed and parallel computing. (ML01 - ML04 - ML05 - DC01).

Awards

2009   First in admission ranking for Ph.D. out of 446 applicants.

2009   Ph.D. scholarship for 3 years sponsored by the European Union.

2008   Award for outstanding academic career from University of Catania.

2006   Erasmus scholarship from ERSU (Regional Agency for University Studies)

2006   Socrates mobility scholarship from Univeristy of Catania.

2005   Master's scholarship from ERSU (Regional Agency for University Studies).

2004   Award for achieving full marks and honors in minimum time from ERSU.

2003   Bachelor's scholarship from ERSU (Regional Agency for University Studies).

2002   Bachelor's scholarship from ERSU (Regional Agency for University Studies).

Languages

Italian	**Native**	
English	**Proficient**	*European level C2*
Spanish	**Proficient**	*European level C2*
Japanese	**Basic**	*European level A2*

Computer skills

Languages	Java, C/C++, Python, Perl, Bash, PHP, SQL, LaTeX.
Big Data	Hadoop ecosystem, Pig, Giraph, S4, Storm.
Tools	Matlab, ANTLR (parser generator), Subversion, git, Ant+Ivy, Maven.

Interests

○ Chess, reading, cooking, RPG, anime, archery, basket, photography.

Patents

Pending

○ Two methods for matching social content in map-reduce architecture.

○ An information retrieval method for online news suggestion using closed captions.

○ An automatic method to build flowcharts for e-shopping recommendations.

Publications

Gianmarco De Francisci Morales and Ajay Shekhawat. The Future of Second Screen Experience. In *Workshop on Exploring and Enhancing the User Experience for Television @CHI '13*, Paris, 2013.

Ilaria Bordino, Gianmarco De Francisci Morales, Ingmar Weber, and Francesco Bonchi. From Machu_Picchu to "rafting the urubamba river" : Anticipating information needs via the Entity-Query Graph. In *WSDM '13: 6th International Conference on Web Search and Data Mining*, pages 275–284, 2013.

Gianmarco De Francisci Morales, Aristides Gionis, and Claudio Lucchese. From Chatter to Headlines: Harnessing the Real-Time Web for Personalized News Recommendation. In *WSDM '12: 5th International Conference on Web Search and Data Mining*, pages 153–162, New York, New York, USA, February 2012.

Gianmarco De Francisci Morales, Aristides Gionis, and Mauro Sozio. Social Content Matching in MapReduce. *VLDB Endowment*, 4(7):460–469, 2011.

Gianmarco De Francisci Morales, Claudio Lucchese, and Ranieri Baraglia. Scaling Out All Pairs Similarity Search with MapReduce. In *LSDS-IR '10: 8th Workshop on Large-Scale Distributed Systems for Information Retrieval @SIGIR '10*, pages 25–30, 2010.

Gianmarco De Francisci Morales. Cloud computing for large scale data analysis. Technical report, IMT Institute for Advanced Studies, Lucca, February 2010.

Ranieri Baraglia, Claudio Lucchese, and Gianmarco De Francisci Morales. Large-scale Data Analysis on the Cloud. In *XXIV Convegno Annuale del CMG-Italia*, Roma, 2010.

Ranieri Baraglia, Gianmarco De Francisci Morales, and Claudio Lucchese. Document Similarity Self-Join with MapReduce. In *ICDM '10: 10th International Conference on Data Mining*, pages 731–736, 2010.

Vincenzo Catania, Alessandro G. Di Nuovo, Maurizio Palesi, Davide Patti, and Gianmarco De Francisci Morales. An Effective Methodology to Multi-objective Design of Application Domain-specific Embedded Architectures. In *DSD '09: 12th Conference on Digital System Design, Architectures, Methods and Tools*, pages 643–650, 2009.

Vincenzo Catania, Gianmarco De Francisci Morales, Alessandro G. Di Nuovo, Maurizio Palesi, and Davide Patti. High Performance Computing for Embedded System Design: A Case Study. In *DSD '08: 11th Conference on Digital System Design Architectures, Methods and Tools*, pages 656–659, 2008.