

# Giovanni Simonini, PhD

## CURRICULUM VITAE

### PERSONAL INFORMATION

Address	Via Vivarelli 10, Modena 41125, Italy
email	<a href="mailto:giovanni.simonini@unimore.it">giovanni.simonini@unimore.it</a>
website	<a href="http://dbggroup.unimo.it/simonini">dbggroup.unimo.it/simonini</a>
phone	(M) +39 3336709437
Skype	<a href="https://www.skype.com/people/giovanni.simonini">giovanni.simonini</a>

### RESEARCH EXPERIENCE

2016-present  
Postdoc at  
UniMoRe

My research activity has been focused on the modelling of novel approaches for data integration techniques in the context of Big Data and Cloud Computing. Currently I am working on devising new methods for scaling Entity Resolution, and developing them on Apache Spark.

Advisor: Prof. Sonia BERGAMASCHI · Università di Modena e Reggio Emilia

2013–2016  
PhD Student at  
UniMoRe

My research activity as PhD student of the DBGroup of the Università di Modena e Reggio Emilia has been focused on the modelling of a new approach for **Entity Resolution (a.k.a. Record Linkage or Duplicate Detection) for Big Data and highly heterogeneous data**. In particular, the novel contribution of my work is related to the modelling and developing of a new technique for blocking (i.e. indexing). During the research activities the following technologies and frameworks has been employed:

**Java; Scala; Python; Apache Spark; AWS; Docker; Lucene; Weka.**

Advisor: Prof. Sonia BERGAMASCHI · Università di Modena e Reggio Emilia

Co-advisor: Prof. H.V. JAGADISH · University of Michigan

### WORK EXPERIENCE

2013–Present  
Collaborator at  
DataRiver

Consultant for Data Integration projects and Supervisor for internship projects.  
[datariver.it/company/team/#GiovanniSimonini](http://datariver.it/company/team/#GiovanniSimonini)

01/12– 12/12  
Software Engineer  
at DataRiver

I won a grant co-financed by Spinner 2013 and DataRiver for developing a cross-platform app for smartphone, and its backend, for enterprise sale departments.

### VISITING & COLLABORATIONS

10/17  
Visiting Research  
at HPI, Potsdam

Collaboration with Prof. Felix Naumann at the Hasso Plattner Institute, Potsdam, on a **Big Data Exploration** project.

03/17– 04/17  
Visiting Research  
at Paris Descartes  
University

Collaboration with Prof. Themis Palpanas at the Descartes University, Paris, on a **Big Data Cleaning** project.

05/15– 11/15  
Visiting Research  
at U. of Michigan

Collaboration with Prof. H.V. Jagadish at the University of Michigan, Ann Arbor, on a **Big Data Integration** project.

### AWARDS AND GRANTS

Awards

2017 · IEEE CS Italy Section Chapter 2016 PhD Thesis Award

Grants

2016 · VLDB travel grant  
2015 · Grant from CINECA - Project: Entity Resolution for Big Data  
2013 · Three year Grant from MIUR for PhD program  
2012 · Spinner 2013 Scholarship

## TEACHING ACTIVITY

2017 UniMoRe	<b>Adjunct Professor</b> · 54 hours front teaching for the course of Foundation of Databases and Information Systems — Sistemi informativi e basi di dati, DISMI, corso di laurea in Ingegneria Gestionale
2013-present UniMoRe	Teaching Assistant · Advanced Data Base Technologies (Tecnologia delle Basi di Dati) - Teaching Assistant
2013-present UniMoRe CINECA	Teaching Assistant · Data Base Foundation (Basi di Dati)  2014-present · Tools and techniques for massive data analysis @CINECA I held three days courses (×4 editions) on tools and techniques for Big Data (focused on the <b>Apache Spark</b> ecosystem), promoted and hosted by <b>CINECA</b> . Link: <a href="#">Tools and techniques for massive data analysis</a>

## EDUCATION

2013–2016 PhD Student at UniMoRe	UNIVERTISTÀ DI MODENA E REGGIO EMILIA ICT INTERNATIONAL DOCTORAL SCHOOL <b>Grade:</b> Ottimo (highest grade with honors) <b>Thesis:</b> <i>Loosely Schema-aware Techniques for Big Data Integration</i> <b>Advisor:</b> Prof. Sonia BERGAMASCHI · Università di Modena e Reggio Emilia <b>Co-advisor:</b> Prof. H.V. JAGADISH · University of Michigan
2009–2011 MS in computer engineering	UNIVERTISTÀ DI MODENA E REGGIO EMILIA <b>Grade:</b> 110/110 cum laude <b>Thesis:</b> <i>Design and Development of an Application for Mobile Devices based on the iOS operating system to support management and planning</i> <b>Description:</b> developing a cross-platform app for smartphone, and its backend, for the enterprise sales department. <b>Advisors:</b> Prof. Sonia BERGAMASCHI and Prof. Riccardo MARTOGLIA
2004–2008 BS in electronic engineering	UNIVERTISTÀ DI MODENA E REGGIO EMILIA <b>Description:</b> Optimisation of the microchannel plate system for spectroscopy measurement.

## COMPUTER SKILLS

Programming Languages	PYTHON; SCALA; JAVA; JAVASCRIPT; SHELL; SQL
Big Data Frameworks	SPARK (SPARK, SPARKSQL, MLlib); HADOOP MAPBLUEUCE AND GIRAPH, LUCENE
Other Frameworks	DOCKER; IPYTHON; APACHE ZEPPELIN; AWS EC2; WEKA; SCIKIT-LEARN; NETWORK-X; GEPHI; SCRAPY; MONGODB; SQL SERVER; MYSQL; SQLITE; L <sup>A</sup> T <sub>E</sub> X; APPCELERATOR TITANIUM

## PUBLICATIONS

- **SOPJ: A Scalable Online Provenance Join for Data Integration.** - IEEE International Conference on High Performance Computing & Simulation, HPCS 2017. - S. ZHU, G. FIAMENI, G. SIMONINI, S. BERGAMASCHI. [\[PDF\]](#)
- **BigBench workload executed by using Apache Flink** - Procedia Manufacturing 11 (2017): 695-702. - S. BERGAMASCHI, L. GAGLIARDELLI, G. SIMONINI, S. ZHU [\[PDF\]](#)
- **BLAST: a loosely schema-aware meta-blocking approach for entity resolution** - PVLDB 9.12 (2016): 1173-1184. - G. SIMONINI, S. BERGAMASCHI, H. V. JAGADISH [\[PDF\]](#) [\[CODE\]](#)

- **Providing Insight into Data Source Topics** - Journal on Data Semantics (2016): 1-18. - G. SIMONINI, F. GUERRA, S. BERGAMASCHI, D. FERRARI, Y. VELEGRAKIS [PDF]
- **Big data exploration with faceted browsing** - IEEE International Conference on High Performance Computing & Simulation, HPCS 2015 - G. SIMONINI, Z. SONG [PDF]
- **Discovering the topics of a data source: a statistical approach** - SWSD Workshop @ISWC 2014 - G. SIMONINI, F. GUERRA, S. BERGAMASCHI [PDF]
- **Towards Declarative Imperative Data-parallel Systems** - 22nd Italian Symposium on Advanced Database Systems, SEBD 2014 - M. INTERLANDI, G. SIMONINI [PDF]
- **Using big data to support automatic Word Sense Disambiguation** - IEEE International Conference on High Performance Computing & Simulation, HPCS 2014 - G. SIMONINI, F. GUERRA [PDF]
- **Supporting Image Search with Tag Clouds: a Preliminary Approach** - Advances in Multimedia, 2015 - G. SIMONINI, F. GUERRA, M. VINCINI [PDF]
- **Keyword Search over Relational Databases: Issues, Approaches and Open Challenges** - Bridging Between Information Retrieval and Databases, 54-73. - S. BERGAMASCHI, F. GUERRA, G. SIMONINI [PDF]

#### OTHER INFORMATION

<i>Courses and Specialization</i>	2013 · Cloudera Hadoop developer training, CGI, Bologna 2014 · Advanced Statistics and Data Mining Summer School. Universidad Politecnica de Madrid
<i>Other Courses</i>	· Functional Programming Principles in Scala - Coursera Verified Certificates · Introduction to Data Science - Coursera · Introduction to Machine Learning - Coursera
<i>Languages</i>	ITALIAN · Mothertongue ENGLISH · Professional
<i>Interests</i>	Running · Squash · Snowboard · MENSA society member

*Autorizzo il trattamento dei dati personali contenuti nel mio curriculum vitae in base art. 13 del D. Lgs. 196/2003.*

December 29, 2017