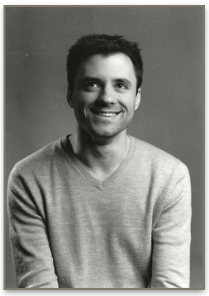
Anthony Simonet, PhD
Born February 15th, 1987 in France
87 rue Paul Bert,
69003 Lyon
France
Phone +33 (0)6 85 42 56 17
anthony.simonet@gmail.com
https://www.anthony-simonet.fr



COMPUTER SCIENCE RESEARCHER

Distributed Computing and Data Management,
Cloud Computing, High-Performance Computing

EDUCATION

- **2011–2015** Ph.D. in Computer Science at École Normale Supérieure de Lyon under the guidance of Gilles Fedak, defended on July 8th, 2015 (mention "Très Honorable").

  Subject: "Active Data: Enabling Smart Data Life Cycle Management for Large Distributed Scientific Data Sets"

  1 month internship at the *Computation Institute* of the *University of Chicago* with Ian Foster that lead to the publication of a short paper in *PDP 2015*.

- **2009–2011** Master's Degree in Computer Science, specialty in Software Engineering and Distributed Software Architecture at the University Bordeaux I (mention "Bien").

  Master's thesis: "A Programming Model for Active Data in Hybrid DCIs" supervised by Gilles Fedak.

  Exchange student (one quarter) at the University of California, San Diego in 2010. Independent Study under the supervision of Tajana Simunic Rosing, head of the *System Energy Efficiency Lab*.

- **2009** Laureate of the Isabelle Attali Prize from University Bordeaux I (http://prix-isabelle-attali.u-bordeaux1.fr/)

- **2007–2009** Licence of Computer Science at University Bordeaux I (mention "Assez Bien")

- **2005–2007** Technical University Degree (DUT) in Statistics and Data Processing (STID) at the University of Pau, France.

EXPERIENCE

- **September 2017–present** Postdoctoral Associate with the Rutgers Discovery Informatics Institute (RDI2) at Rutgers University (New Jersey, USA) with Pr. Manish Parashar.

- **October 2015–June 2017** Postdoctoral Researcher at Inria and Institut Mines Télécom Atlantique (Nantes, France) with Pr. Adrien Lebre.

MAJOR PROJECTS

- GeoSCIFramework (Rutgers University, ongoing): NSF-funded project aiming at building an interactive platform for facilitating the distribution of data produced by geophysics sensor networks, and the development of geoscience applications. I was involved in writing the project proposal (funded by the American National Science Foundation for $800,000 starting January 2019) and I am working on a use-case for Tsunami Early Warning. For

this application we leverage Machine Learning techniques to process the time series produced by different sensors at the edge and *in-transit*, in order to improve the precision of alerts and reduce the delays to issue them.

- Virtual Data Collaboratory (VDC) (Rutgers University, ongoing): NSF-funded project aiming at producing a blueprint and building a regional cyber-infrastructure for collaborative and inter-disciplinary research. I am mentoring a 4th year PhD student who is developing an Internet-scale caching and staging service that will be deployed in the participating universities. This framework will accelerate data exchanges between tasks and facilitate application development by offering high-level APIs with simple *put/get* abstractions. I am also working on the software architecture (management of servers and tasks, computation services for users).
- DataSpaces (Rutgers University, ongoing): in-memory staging software for HPC developed at RDI2. I am mentoring a 2nd year PhD student working on a novel event-based scientific workflow management system. His prototype will eventually allow users to specify predicates that DataSpaces will match against the intermediate data produced by workflow tasks, in order to determine the execution of the next tasks.
- Outreach activities (Rutgers University, ongoing): development of the "Diving into Big Data" workshop for middle- and high-schools in New Jersey. We welcome groups of about 20 students and give them an introduction to research in general and to some of the research taking place at the department. Students visit the Ocean Observatory Initiative (OOI) data center and take part in a hands-on session, during which they program a simple dashboard displaying and transforming real-time ocean data streaming from OOI.
- Discovery Initiative (Inria, 2015−2017): collaboration between several Inria research groups in France aiming at implementing an Edge Computing platform that follows the IaaS cloud computing model. In this context, I investigated the feasibility of the project from the viewpoint of energy usage and cost. I co-mentored an intern as she developed tools for measuring the performance of OpenStack on multiple infrastructures.

Teaching

I taught approximately 64 hours a year at the University Lyon 1 during my Ph.D. in 2012 and 2013:

- Practical Sessions Algorithms/Functional Programming, 1st year students: 28 hours
- Practical Sessions Parallelism, 4th year students: 12.5 hours
- Group Work & Practical Sessions Operating Systems, 3rd year students: 44 hours
- Practical Sessions Parallelism, 4th year students : 12.5 hours

I taught a total of 27 hours at the Institut Mines-Télécom Atlantique (engineering program) from 2015 to 2017:

- Lectures & Practical Sessions Cloud Computing (AWS), 4th year engineering students: 18 hours (2015 & 2016)
- Practical Sessions MapReduce (Hadoop), 4th year engineering students: 9 hours (2016 & 2017)

PUBLICATIONS

➡ *International journals*

- ***The Virtual Data Collaboratory: A Regional Cyberinfrastructure for Collaborative Data-Driven Research*** – Manish Parashar, Vasant Honavar, Anthony Simonet, Ivan Rodero, Forough Ghahramani, Grace Agnew, Ron Jantz, to appear in IEEE Computing in Science Engineering.
- ***Towards a Computing Continuum: Enabling Edge-to-Cloud Integration for Data-Driven Workflows*** – Daniel Balouek-Thomert, Eduard Gibert Renard, Ali Reza Zamani, Anthony Simonet, Manish Parashar, to appear in Special Issue of International Journal of High Performance Computing Applications.
- ***Putting the Next 500 VM Placement Algorithms to the Acid Test*** – Adrien Lebre, Jonathan Pastor, Anthony Simonet, Mario Südholt, IEEE Transactions on Parallel and Distributed Systems, Volume 30, Number 1, p. 204-217, 2019.
- ***Active Data: A Programming Model to Manage Data Life Cycle Across Heterogeneous Systems and Infrastructures*** – Anthony Simonet, Gilles Fedak, Matei Ripeanu, Future Generation Computer Systems journal, p. 25-42 – 2015.
- ***Scalable Data Management for Map–Reduce–Based Data–Intensive Applications: a View for Cloud and Hybrid Infrastructures*** – Gabriel Antoniu, Alexandru Costan, Julien Bigot, Frédéric Desprez, Gilles Fedak, Sylvain Gault, Christian Pérez, Anthony Simonet, Bing Tang, Christophe Blanchet, Raphael Terreux, Luc Bougé, François Briant, Franck Cappello, Kate Keahey, Bogdan Nicolae, Frédéric Suter, International Journal of Cloud Computing, Volume 2, Number 2, p. 150-170 – 2013

➡ *International conferences*

- ***Towards a Holistic Framework for Conducting Scientific Evaluations of OpenStack*** – Ronan-Alexandre Cherrueau, Adrien Lebre, Dimitri Pertin, Anthony Simonet, Matthieu Simonin, Short paper in Proceeding of the IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGrid 2017 – IEEE, 2017
- ***Revising OpenStack to Operate Fog/Edge Computing Infrastructures*** – Adrien Lebre, Jonathan Pastor, Anthony Simonet, Frédéric Desprez, Proceedings of the IEEE International Conference on Cloud Engineering, IC2E 2017 – IEEE, 2017
- ***Deploying Distributed Cloud Infrastructures: Who and at What Cost?*** – Anthony Simonet, Adrien Lebre, Anne-Cécile Orgerie, Proceedings of the 5th IEEE International Workshop on Cloud Computing Interclouds, Multiclouds, Federations and Interoperability, Intercloud 2016 – IEEE, 2016
- ***D³ MapReduce: Towards MapReduce for Distributed and Dynamic Data Sets*** – Heithem Abbes, Julio Anjos, Asma Ben Cheikh, Gheorghe Cosmin Silaghi, Gilles Fedak, Haiwe He, Hai Jin, Lu Lu, Mircea Moca, José-Francisco Saray, Xuanhua Shi, Anthony Simonet, Bing Tang, Proceedings of the 2015 International Conference on Big Data Intelligence, DataCom 2015 – IEEE, 2015
- ***Using Active Data to Provide Smart Data Surveillance to E-Science Users*** – Anthony Simonet, Gilles Fedak, Kyle Chard, Ian Foster, Proceedings of the 23rd Euromicro International Conference on Parallel, Distributed and Network-based Processing conference, PDP 2015 (Short paper) – IEEE, 2015
- ***Active Data: a Data-Centric Approach to Data Life-Cycle Management*** – Anthony Simonet, Gilles Fedak, Matei Ripeanu, Samer Al-Kiswany, Proceedings of the 8th Parallel Data Storage Workshop – ACM, 2013
- ***Towards Scalable Data Management for Map-Reduce-based Data-Intensive Applications on Cloud and Hybrid Infrastructures*** – Gabriel Antoniu, Julien Bigot, Christophe Blanchet, Luc Bougé, François Briant, Franck Cappello, Alexandru Costan, Frédéric Desprez, Gilles Fedak, Sylvain Gault, Kate Keahey, Bogdan Nicolae, Christian Pérez, Anthony Simonet, Frédéric Suter, Bing Tang, Raphael Terreux, 1st International IBM Cloud Academy Conference, ICA CON 2012

➡ *Others*

- ***Energy-Aware Massively Distributed Cloud Facilities: the DISCOVERY Initiative***, Frédéric Deprez, Shadi Ibrahim, Adrien Lèbre, Anne-Cécile Orgerie, Jonathan Pastor, Anthony Simonet, Poster, 11th IEEE International Conference on Green Computing and Communications, GreenCom 2015
- ***Active Data : Un modèle pour représenter et programmer le cycle de vie des données distribuées*** – Anthony Simonet, ComPAS'2014, nationale conference, peer-reviewed
- ***MapReduce on Desktop Grids with BitDew and Active Data*** – Anthony Simonet, Lu Lu, Xuanhua Shi, Bing Tang, José-Francisco Saray, Gilles Fedak, Grid'5000 Winter School 2012
- ***Active Data: A Programming Model to Manage Data Life Cycle Across Heterogeneous Systems and Infrastructures*** – Anthony Simonet, Gilles Fedak, Matei Ripeanu, research report, 2013
- ***FlyingGrid: from Volunteer Computing to Volunteer Cloud*** – Oleg Lodygensky, Etienne Urbah, Simon Dadoun, Gilles Fedak, Simon Delamare, Derrick Kondo, Laurent Duflot, Xavier Garrido, Anthony Simonet, Peter Kacsuk, Jozsef Kovacs, Zoltan Farkas, poster in Computing in Hign Energy and Nuclear Physics CHEP'12, 2012, poster

➡ *Ongoing Work*

- ***Towards a Smart Internet-Scale Cache and Staging Service for Data-Intensive Scientific Applications*** – Yubo Qin, Anthony Simonet, Ivan Rodero, Manish Parashar
- ***Leveraging Scalable Event Distribution for Orchestrating Data-driven In-situ Scientific Workflows*** – Zhe Wang, Anthony Simonet, Pradeep Subedi, Philip E. Davis, Manish Parashar
- ***Towards a Distributed Tsunami Early-Warning System based on Multi-sensor Machine Learning Techniques*** – Kevin Fauvel, Anthony Simonet, Daniel Balouek-Thomert, Pedro Silva, Ivan Rodero, Manish Parashar, Gabriel Antoniu

Communications and Talks

- Seminar at the *Department of Computer Science* of Columbia University, New York City, 2012
- Presentation at the *MapReduce ANR* Meeting, IBM France, Montpellier, France, 2012
- Seminar at the *Computation Institute* of the *University of Chicago*, 2013
- Presentation at the *Joint Lab on Petascale Computing* workshop, Inria Sophia Antipolis, France, 2014
- Seminar at the *DB* research group of the *LIRIS* at the *Université Lyon 1*, Lyon, France, 2015
- Seminar at the *Discovery* Plenary Meeting, Nantes, France, 2015
- Talk at the *Large Scale Deployment Challenge* of the Grid'5000 Winter School, Grenoble, 2016
- Talk at the *Journée OpenStack Workshop* in Lyon, France, 2016
- Talk at the *Cloud Control Workshop*, Uppsala, Sweden, 2016
- Seminar at the *Discovery* Plenary Meeting, Rennes, France, 2016
- Talk at the *Journée OpenStack Day France*, Paris, 2016
- Seminar at the *Avalon* research group of the *LIP* at the *École Normale Supérieure de Lyon*, Lyon, France, 2017
- Invited talk at *NJEdgeCon 2018*, Whyppany, New Jersey, 2018

Chair/Program Committee

- PC member at the *10th IEEE International Conference on Cloud Computing Technology and Science* (CloudCom 2018)
- PC member at the *IEEE International Conference on High Performance Computing and Communications* (HPCC 2018)
- PC member at *the IEEE International Congress on Internet of Things* (ICIOT 2018)
- PC member at the *IEEE International Congress on Big Data* (BigData 2018)
- PC member at the *9th International Conference on Cloud Computing Technology and Science* (CloudCom 2017)
- PC member at the *13th International Conference on Green Computing and Communications* (GreenCom 2017)
- PC member at the *14th International Symposium on Pervasive Systems, Algorithms, and Networks* (I-SPAN 2017)
- PC Chair at the *1st IEEE International Conference on Fog and Edge Computing* (ICFEC'2017)
- PC member of the 1*2th International Conference on Economics of Grids, Clouds, Systems and Services* (GECON 2015)

Reviewer

- IEEE Cloud Computing Magazine (IEEE Cloud Computing)
- IEEE Transactions on Cloud Computing (IEEE TCC)
- IEEE Transactions on Parallel and Distributed Systems (IEEE TPDS)
- IEEE Transactions on Services Computing (IEEE TSC)
- ACM Transactions on Internet Technology (ACM TOIT)
- 27th ACM International Symposium on High Performance Parallel and Distributed Computing (HPDC 2018)
- 38th IEEE International Conference on Distributed Computing systems (ICDCS 2018)
- 47th International Conference on Parallel Processing (ICPP 2018)
- International European Conference on Parallel and Distributed Computing (Europar 2016)
- International Symposium on Stabilization, Safety and Security of Distributed Systems (SSS 2016)
- International Workshop on Optimization of Energy Efficient HPC & Distributed Systems (OPTIM 2016)
- International Workshop on High Performance Data Intensive Computing (HPDIC 2013)
- IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2012)
- IEEE International Conference on e-Science (e-Science 2012)
- International Workshop on MapReduce and its Applications (MapReduce 2012)