Jad DARROUS | Curriculum Vitae

15 Rue du Gros Chêne, 44300 Nantes, France

+33 7 82.97.26.89 • jad.darrous@gmail.com • jad-darrous.github.io

Education

Academic Qualifications

- **Ph.D. in Computer Science: Large-scale Data Management Systems** — **Lyon, France**
  École Normale Supérieure (ENS) de Lyon — *2016 – 2019*
- **Master in Computer Science: Parallel and Distributed Systems** — **Grenoble, France**
  Grenoble-Alpes University — *2014 – 2016*
- **Bachelor in Computer Science** — **Damascus, Syria**
  Damascus University — *2007 – 2012*

Experience

- **Inria Rennes Bretagne-Atlantique** — **Nantes, France**
  Research Engineer — *Jan 2020 – Aug 2020*
  Design and develop dynamic erasure coding chunk scheduling in Hadoop file system (HDFS).
  - Implement a StripeReader that can read the original data from any n out of $n + k$ chunks at the client side.
  - Implement a dynamic load balance unit at the NameNode that returns to the HDFS client the best locations to read data from.

- **Inria Grenoble Rhône-Alpes** — **Lyon, France**
  Research student — *Oct 2016 – Dec 2019*
  Conducted research on scalable and efficient data management in distributed clouds: Service provisioning and Big Data processing. Contributions:
  - Design, develop and implement NITRO, a Virtual Machine Image (VMI) management system in geo-distributed clouds. NITRO leverages data deduplication and employs a flow-based optimal scheduling algorithm to reduce network cost and improve VM provisioning time by up to 77% compared to OpenStack Swift. **Codebase**: 1500 LoC in Python. **Testbed**: 12 physical machines plus network emulation. **Tools**: Redis, BitTorrent, OpenStack Swift, zeromq, networkx, json, pandas, matplotlib, Ansible, Docker, Vagrant.
  - Design and develop a simulator for container image placement algorithms in Edge environment. Based on k-Center optimization, we implement two placement algorithms that can reduce the maximal retrieval time for container provisioning by up to 4x compared to state-of-the-art algorithms. **Codebase**: 2500 LoC in Python. **Dataset**: IBM container traces. **Tools**: PyPy, CPLEX, json, yaml, pandas, matplotlib.
  - Perform an in-depth experimental evaluation of Big Data applications running under *erasure coding* (EC). **Testbed**: 61 physical machines (21 Hadoop cluster + 40 clients). **Tools**: Hadoop, MapReduce, Yarn, Bash, psutil, pandas, matplotlib. **Experiments**: more than 1000 experiments, fully automated.
  - Design and develop data placement algorithm to improve EC data access performance in HDFS. The algorithm can reduce access time by up to 34% under concurrent clients' reads. **Implementation**: 200 LoC in Java at the NameNode in HDFS.

- **Syrian Educational Publisher Co., for E-Publishing and E-learning** — **Damascus, Syria**
  Java Developer — *Jan 2013 – Aug 2014*
  Accomplished tasks, while working in Scrum environment:
  - Extend the core libraries to cover more of the EPUB3 spec.
  - Implement unit tests for the core libraries with code coverage of 70%.
  - Develop 60% of the Android application (Epub3 Viewer).
  - Implement the first prototype of the GWT (Google Web Toolkit) website.
  - Implements synchronization between applications running on different devices.

Technical skills

- Programming languages: Python, Java, and C++.
- Frameworks and tools: Hadoop, Spark, Docker, OpenStack, Ansible, Vagrant, and Git.
- Academic experience with (SQL and NoSQL) databases.
- Academic experience in Natural Language Processing, Computer Vision and Machine Learning.
- Familiar with mobile (android) and web (front-end/back-end) technologies.
- Critical thinking, creative problem-solving, teamwork spirit, and writing skills.

Publications

Papers in international conferences

- **Jad Darrous**, Shadi Ibrahim, Christian Perez, "Is it time to revisit Erasure Coding in Data-intensive clusters?" in *Proceedings of the* 27^{th} *IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)* - Oct. 2019, Rennes, France.
- **Jad Darrous**, Thomas Lambert, Shadi Ibrahim, "On the Importance of Container Image Placement for Service Provisioning in the Edge" in *Proceedings of the* 28^{th} *International Conference on Computer Communications and Networks (ICCCN)* - Jul. 2019, Valencia, Spain.
- **Jad Darrous**, Shadi Ibrahim, Amelie Chi Zhou, Christian Perez, "Nitro: Network-Aware Virtual Machine Images Management in Geo-Distributed Clouds" in *Proceedings of the* 18^{th} *IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)* - May. 2018, Washington DC, USA.

Posters in international conferences

- **Jad Darrous**, Shadi Ibrahim, "Enabling Data Processing under Erasure Coding in the Fog" in *The* 48^{th} *International Conference on Parallel Processing (ICPP)* - Aug. 2019, Kyoto, Japan.

Honors and Awards

- PERSIVAL-Lab scholarship for academic excellence for the year 2015-2016.
- I was an active member in programming competitions (2009-2014):
  - I represented Damascus University (Faculty of Informatics) in the ACPC programming contest - the regional tier of ICPC - within a team of 3 students, and we got the 8^{th}, 14^{th}, and 11^{th} place consecutively from 2009 to 2011.
  - I coached the teams of Damascus University in the ACPC programming contest 2012 and 2013.
  - I was a judge in the Syrian Collegiate Programming Contest - the national tier of ICPC - in 2012 and 2013.
  - I was a problem setter in many national and local contests.
- The 8^{th} place in the fourth annual *Math Olympiad* contest held by Department of mathematics, Damascus University 2012.
- We won, as a team of three, the 3^{rd} place in *Fikra Competition* for our "Simple Geometry Solver" project, Damascus 2011.

Languages

- Professional proficiency in French and English.
- Native Arabic speaker.