Tu Mai Anh Do

https://tumaianhdo.github.io

Updated: September, 2022
Email: tumaianhdo@gmail.com
Mobile: (385) 528-8701

Research Interests

Big Data Analytics, Performance Modeling, Resource Scheduling, Workflow Management, High Performance Computing, Distributed Systems, Machine Learning

Education

- **University of Southern California (USC)** — Los Angeles, California, USA
  Ph.D. Candidate in Computer Science — 2017 - Present
  - Relevant Coursework: High Performance Computing and Simulations, Advanced Topics in Database Systems, Foundations and Applications of Data Mining, Machine Learning, Foundations of Optimization
  - Advisor: Ewa Deelman
- **Ho Chi Minh City University of Technology (HCMUT)** — Ho Chi Minh, Vietnam
  Bachelor of Engineering in Computer Engineering, Honors Program — GPA: 8.55/10.00 — 2011 - 2016
  - Thesis: Developing Methods To Help Large-Scale Parallel Applications More Reliable – Highest score among concurrent defensed theses

Research Experience

- **Information Sciences Institute, Science Automation Technologies** — Marina Del Rey, California
  Research Assistant — 2017 - Present
  - Introduced computational efficiency model to quantify efficiency of in situ workflows, a processing paradigm that provides ability to analyze data as it is generated and store only necessary data (published in JOCS 2021)
  - Proposed multi-stage performance indicators that capture performance of entire in situ workflow ensembles in terms of multiple resource perspectives (published in CCPE 2022)
  - Determined co-scheduling strategies and resource assignments for in situ workflow ensembles such that makespan is minimized (accepted to WORKS 2022)
- **Lawrence Livermore National Laboratory, Center for Applied Scientific Computing** — Livermore, California
  Research Intern — 2018
  - Enabled scientific workflows that couple high-performance simulations with big data analytics by leveraging node-local storage to reduce expensive storage needs for storing large datasets (presented in SC 2018 research poster)
- **High Performance Computing Laboratory, HCMUT** — Ho Chi Minh, Vietnam
  Research Assistant — 2014 - 2017
  - Developed techniques to detect abnormal behaviors, e.g. message leak, race condition, deadlock, for large-scale parallel applications using message-passing programming model (published in ISPDC 2016)

Notable Projects

- **In Situ Data Analytics for Next Generation Molecular Dynamics Workflows** — 2017 - 2022
  - Designed a runtime that allows to decouple in situ analyses from the simulation and address decoupling complexities in terms of data coupling incompatibility
- **Yelp Recommendation based on Aggregated Ratings** — 2019
  - Combined sentiment score from text mining analysis to normalize star ratings, which then used in a collaborative filtering recommendation system
- **Sale Forcasting for Kaggle's Predict Future Sales** — 2020
  - Applied Long-Short Term Memory (LSTM) and an ensemble of independent Extreme Gradient Boosting (XGBoost) models to improve precision of monthly sales prediction
- **Accelerating Scientific Workflows on HPC Platforms with In Situ Processing** — 2020 - 2021
  - Integrated in situ technology with traditional workflow management system through clustering data-intensive jobs (published in CCGrid 2022)
- **Quantum Acceleration for Scientific Computing** — 2021 - 2022
  - Provided a formal method for the selection of hyperparameters in designing variational quantum algorithms for scientific applications (accepted to eScience 2022)

Work Experience

- **Novobi** — Ho Chi Minh, Vietnam
  Software Engineer — Mar 2017 - July 2017
  - Built an automated system for deploying, testing and delivering software packages of health care applications
- **DEK Technologies** — Ho Chi Minh, Vietnam
  Software Engineer Intern — May 2015 - Aug 2015
  - Automated the deployment of small-scale clusters with high availability

Teaching Experience

- **University of Southern California** — Los Angeles, California, USA
  Teaching Assistant — Fall 2020, Spring 2021
  - Courses: Database Systems (CSCI 585, Master-level)
  - Assisted students with course material and their programming assignments, graded midterm and final exams
- **Ho Chi Minh City University of Technology** — Ho Chi Minh, Vietnam
  Teaching Assistant — Sep 2016 - May 2017
  - Courses: Parallel Programming and Distributed Systems, Fundamentals of Programming
  - Held lab sessions and assisted students with their programming assignments

Selected Publications

- **Do, T. M. A.**, Pottier, Ferreira da Silva, R., L., Caíno-Lores, S., T., Taufer, M., and Deelman, E. *Performance assessment of ensembles of in situ workflows under resource constraints.* Concurrency and Computation Practice and Experience (CCPE), 2022
- **Do, T. M. A.**, Pottier, L., Yildiz, O., Vahi, K., Krawczuk, P., Peterka, T., and Deelman, E.. *Accelerating Scientific Workflows on HPC Platforms with In Situ Processing.* IEEE/ACM 22nd International Symposium on Cluster, Cloud and Internet Computing (CCGrid), 2022.
- **Do, T. M. A.**, Pottier, L., Caíno-Lores, S., Ferreira da Silva, R., Cuendet, M. A., Weinstein, H., Estrada, T., Taufer, M., and Deelman, E.. *A Lightweight Method for Evaluating In Situ Workflow Efficiency.* Journal of Computational Science (JOCS), 2021
- Ferreira da Silva, R., Callaghan, S., **Do, T. M. A.**, Papadimitriou, G., and Deelman, E.. *Measuring the Impact of Burst Buffers on Data-Intensive Scientific Workflows.* Future Generation Computer Systems (FGCS), 2019
- Thomas, S., Wyatt, M., **Do, T. M. A.**, Pottier, L., Ferreira da Silva, R., Weinstein, H., Cuendet, M. A., Estrada, T., Deelman, E., and Taufer, M. *Characterization of In Situ and In Transit Analytics of Molecular Dynamics Simulations for Next-generation Supercomputers.* 15th International Conference on eScience (eScience), 2019
- **Do, T. M. A.**, Diep, T., and Thoai, N. *Race Condition and Deadlock Detection for Large-Scale Applications.* 15th International Symposium on Parallel and Distributed Computing (ISPDC), 2016

Honors and Awards

- ISI Distinguished Top-Off Fellowship — 2017
- Elected Candidate of Vietnam Education Foundation (VEF) Fellowship — 2016
- 18th Eureka Scientific Research Student Award Finalist — 2016
- 7th HCMC Information and Communication Technology Award for Student — 2015
- DATALOGIC Vietnam's Scholarship, CSC Vietnam's Scholarship — 2014

Professional Services

- Sub-reviewer for International Conference on Parallel Processing (ICPP) — 2018, 2022
- Sub-reviewer for International Symposium on Cluster, Cloud and Internet Computing (CCGrid) — 2020
- Reviewer for International Conference on Parallel Processing and Applied Mathematics (PPAM) — 2019
- Student volunteer for Supercomputing (SC) — 2018, 2019, 2021

Online Courses and Certifications

- **Oxford Machine Learning Summer School (OxML)**, ML x Finance Track — 2022

Technical Skills

- **Languages:** C/C++, Python, Bash, Java, Scala, Go
- **Technologies:** Docker, AWS, SQL(MySQL, SQLite), NoSQL(PostgreSQL), Apache Hadoop, Apache Spark
- **Machine Learning:** XGBoost, scikit-learn, PyTorch, Optuna
- **Tools:** Pegasus WMS, Dataspaces, Git, CMake, Visual Studio, LATEX, JIRA, Confluence
- **Operating Systems:** Unix/Linux, MacOS, MS Windows

References

- **Rafael Ferreira da Silva**
  Senior Research Scientist
  Oak Ridge National Laboratory
  Oak Ridge, TN, USA
  ✉silvarf@ornl.gov
- **Loïc Pottier**
  Computer Scientist
  USC Information Sciences Institute
  Marina Del Rey, CA, USA
  ✉lpottier@isi.edu