Tu Mai Anh Do

Contact Information

Email: tudo@isi.edu
Personal Website: https://tumaianhdo.github.io
Address: Room 1220, 4676 Admiralty Way, Suite 1001
Marina del Rey, California 90292, USA

Research Interests

In situ Processing, Big Data Analytics, Scientific Workflow Management, High Performance Computing, Distributed Systems

Education

Aug. 2017 – Now
University of Southern California (USC), Los Angeles, California, USA
Viterbi School of Engineering
Ph.D. Candidate in Computer Science
Advisor: Dr. Ewa Deelman

Sep. 2011 – April. 2016
Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh, Vietnam
Faculty of Computer Science and Engineering
B.E. in Computer Engineering, Honors Program
GPA: 8.55/10.00

Experience

Aug. 2017 – Now
Information Sciences Institute, Marina Del Rey, California
Graduate Research Assistant
Advisor: Dr. Ewa Deelman
I am part of the Science Automation Technologies group and mainly working in the project Analytics4MD. My research topic is more concerned about how to enable efficient executions of in situ analyses that co-locates with large-scale simulations using in-memory staging in leadership-class computing platforms.
Tools: DataSpaces, ADIOS, Decaf, Pegasus, Cmake, Shippable, Gromacs, Plumed2, A4MD, TAU

May. 2018 – Aug. 2018
Lawrence Livermore National Laboratory, Livermore, California
Student Intern
Advisor: Dr. Ming Jiang
I collaborated with computer scientists in the Center for Applied Scientific Computing (CASC) to enable data analytics workflows that couple high-performance computing simulations with Big Data analytics using node-local storage.
Tools: Spark, Ascent, Pegasus

Oct. 2014 – Jul. 2017
High Performance Computing Laboratory, Ho Chi Minh City University of Technology
Research Assistant
Advisor: Dr. Nam Thoai
I worked closely with Thanh-Dang Diep and developed techniques to detect abnormal behavior for large-scal parallel applications in message-passing programming model.
Tools: C, C++, MPI, Makefile, Shell Script

Mar. 2017 – Jul. 2017
Novobi, Ho Chi Minh City
Software Engineer
I worked with Cloud Infrastructure team and built an automated system for deploying, testing and delivering packages of health care applications.
Tools: Batch Script, Powershell Script, Atlassian Bamboo, Visual Studio, SQL Server Management Studio

May. 2015 – August. 2015
DEK Technologies, Ho Chi Minh City
Summer Intern
I proposed a high availability solution for computer clusters and automate the deployment in small-scale clusters.
Tools: Shell Script, Git

Teaching Experience

Fall 2020	University of Southern California (USC), Los Angeles, California, USA
:	*Teaching Assistant*
Spring 2021	DATABASE SYSTEMS, *Instructor*: Dr. Saty Raghavachary

Sep. 2016	Ho Chi Minh City University of Technology, Ho Chi Minh City
:	*Teaching Assistant*
:	PARALLEL PROGRAMMING AND DISTRIBUTED SYSTEMS, Instructor: Dr. Nam Thoai
May. 2017	FUNDAMENTALS OF PROGRAMMING, Instructor: Dr. Sach Le

Notable Projects

- **Analytics4MD**

  Name: In Situ Data Analytics for Next Generation Molecular Dynamics Workflows
  Advisor: Dr. Ewa Deelman
  Description: Molecular dynamics simulations studying the classical time evolution of a molecular system at atomic resolution are widely recognized in the fields of scientific computing. Moving to exascale requires the simulations to analyze data as it is generated and store only necessary data. The analysis need to perform in-situ. We propose a paradigm for transforming the centralized, off-line nature of the molecular dynamics analysis into performing in-situ processing via in-memory staging area or in-transit processing via optimized parallel file systems or emerging burst buffer. This effort provides the ability to interleave simulations and analytics to improve data analyzing performance and processing more data.

- **HCMUT B.E Graduation Thesis**

  Title: Developing Methods To Help Large-Scale Parallel Applications More Reliable
  Advisor: Dr. Nam Thoai
  Description: This thesis explored the idea that scalability is the main cause of suffering high probability of error in large-scale parallel applications. To help the applications more reliable for better performance goal, the technique was proposed by detecting abnormal behaviors, which are considered as highly vulnerable from errors.
  Score: 9.96/10.00 – Highest score among concurrent defensed theses

Professional Services

2020	Reviewer for IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing
2019	Reviewer for International Conference on Parallel Processing and Applied Mathematics
2018	Reviewer for International Conference on Parallel Processing

Activities

- *Will be attending* Supercomputing (SC21) as a Virtual Student Volunteer, November 14th, 2021 – November 19th, 2021
- *Presented* a talk titled "Enabling Efficient Execution of In Situ Workflows" at the University of Tennessee, Knoxville Seminar, September 3rd, 2021
- *Attended* Student Volunteers program of Supercomputing (SC19), November 16th, 2019 - November 22nd, 2019
- *Presented* a poster titled "Enabling Data Analytics Workflows using Node-Local Storage" at the USC ISI Graduate Student Symposium 2019, Marina Del Rey, California, March 26, 2019
- *Attended* Supercomputing (SC18) as a Student Volunteer and *presented* an acceptted poster titled "Enabling Data Analytics Workflows using Node-Local Storage", Nov 10, 2018 - Nov 16, 2018
- *Visited* Rutgers Discovery Informatics Institute (RDI^2), New Brunswick, New Jersey, May 13th, 2018 - May 17th, 2018
- *Presented* a poster titled "In Situ Data Analytics for Next Generation Molecular Dynamics Workflows" at the USC ISI Graduate Student Symposium 2018, Marina Del Rey, California, April 5th, 2018

Awards and Honors

2017 | ISI Distinguished Top-Off Fellowship

2016 | Recommended Candidate of 2017 VEFSTA Fellowship Program

2016 | 18th Eureka Scientific Research Student Award Finalist

2015 | 7th HCMC Information and Communication Technology Award for Student

2014 | DATALOGIC Vietnam's Scholarship, CSC Vietnam's Scholarship

Publications

[1] **T. M. A. Do**, L. Pottier, R. Ferreira da Silva, S. Caíno-Lores, M. Taufer, and E. Deelman. "Assessing Resource Provisioning and Allocation of Ensembles of In Situ Workflows". In: *International Workshop on Parallel Programming Models and Systems Software for High-End Computing (P2S2)*, 2021. Funding Acknowledgments: NSF 1741040, NSF 1741057, DOE DE-SC0012636. DOI: 10.1145/3458744.3474051.

[2] **T. M. A. Do**, L. Pottier, S. Caíno-Lores, R. Ferreira da Silva, M. A. Cuendet, H. Weinstein, T. Estrada, M. Taufer, and E. Deelman. "A Lightweight Method for Evaluating In Situ Workflow Efficiency". In: *Journal of Computational Science*, 48, 101259, 2020. Funding Acknowledgments: NSF 1741040, DOE DE-SC0012636. DOI: 10.1016/j.jocs.2020.101259.

[3] **T. M. A. Do**, L. Pottier, S. Thomas, R. Ferreira da Silva, M. A. Cuendet, H. Weinstein, T. Estrada, M. Taufer, and E. Deelman. "A Novel Metric to Evaluate In Situ Workflows". In: *International Conference on Computational Science (ICCS)*, 2020. Funding Acknowledgments: NSF 1741040. DOI: 10.1007/978-3-030-50371-0_40.

[4] S. Thomas, M. Wyatt, **T. M. A. Do**, L. Pottier, R. Ferreira da Silva, H. Weinstein, M. A. Cuendet, T. Estrada, E. Deelman, and M. Taufer. "Characterization of In Situ and In Transit Analytics of Molecular Dynamics Simulations for Next-generation Supercomputers". In: *15th eScience Conference*, 2019. Funding Acknowledgments: NSF 1741040. DOI: 10.1109/eScience.2019.00027.

[5] R. Ferreira da Silva, S. Callaghan, **T. M. A. Do**, G. Papadimitriou, and E. Deelman. "Measuring the Impact of Burst Buffers on Data-Intensive Scientific Workflows". In: *Future Generation Computer Systems*, vol. 101, p. 208–220, 2019. Funding Acknowledgments: DOE DE-SC0012636, NSF 1664162, NSF 1741040. DOI: 10.1016/j.future.2019.06.016.

[6] **T. M. A. Do**, M. Jiang, B. Gallagher, A. Chu, C. Harrison, K. Vahi, and E. Deelman. "Enabling Data Analytics Workflows using Node-Local Storage". In: The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC18), Poster, 2018. Funding Acknowledgments: LDRD 16-ERD-03, NSF 1741040.

[7] **T. M. A. Do**, D. Diep, and N. Thoai. "Race Condition and Deadlock Detection for Large-Scale Applications". In: *2016 15th International Symposium on Parallel and Distributed Computing (ISPDC)*, 2016. DOI: 10.1109/ISPDC.2016.53.

[8] **T. M. A. Do**, D. Diep, and N. Thoai. "Message Leak Detection in Debugging Large-Scale Parallel Applications". In: *2015 International Conference on Advanced Computing and Applications (ACOMP)*, 2015. DOI: 10.1109/ACOMP.2015.17.