Data Science Seminar Series

Toward a Scalable Computing Ecosystem Advancing Data-Integrated Applications for Science and Society

İlkay Altıntaş, Ph.D.
Chief Data Science Officer
San Diego Supercomputer Center
University of California San Diego

Date: Wednesday, March 31st, 2021
Time: 4:00 PM – 5:00 PM EDT
Location: Zoom Virtual Room
Web Link: https://njit-institute-for-data-science.eventbrite.com

Cyberinfrastructure is everywhere in diverse forms. From IoT to extreme-scale computing. Data and computing have never been so distributed with the potential for real-time integration via fast networking and container management. The growth of big data and new processors over the last decade (e.g., GPUs, FPGAs, and edge accelerators) enabled a diverse set of applications using machine learning and data processing on top of distributed nontraditional systems. The common theme to these applications, mostly composed of (big) data-integrated workloads, is their need to run in specialized environments for reasons such as on-demand or 24x7 nature of the tasks they are performing, and difficulties regarding their portability, latency, privacy, and performance optimization. Moreover, in many data-driven scientific applications, there is a need for heterogeneous integration of tasks requiring specialized computing capabilities with traditional high-throughput computing or high-performance computing tasks. Although some key middleware technologies enabled demonstration of standalone heterogeneous applications, such integration requires expertise from a large group of people in very specialized settings. There are still many challenges for streamlined, scalable, repeatable, responsible, and explainable integration of data-integrated applications. Key opportunities for further innovations include intelligent systems and automated workflow management software that can compose and steer dynamic applications that can adapt to changing conditions in a data-driven fashion while integrating many tools to explore, analyze and utilize data. This talk will discuss some examples for data-integrated applications, describe emerging systems that enabled these applications, and overview our recent research to enable composable applications including an application development methodology, intelligent middleware, and workflow composition.

Dr. İlkay Altıntaş, a research scientist at the University of California San Diego, is the Chief Data Science Officer of the San Diego Supercomputer Center as well as a Founding Fellow of the Halıcıoğlu Data Science Institute. She is the Founding Director of the Workflows for Data Science (WorDS) Center of Excellence and the WIFIRE Lab. The WoRDS Center specializes in the development of methods, cyberinfrastructure, and workflows for computational data science and its translation to practical applications. The WIFIRE Lab is focused on artificial intelligence methods for an all-hazards knowledge cyberinfrastructure, becoming a management layer from the data collection to modeling efforts, and has achieved significant success in helping to manage wildfires. Since joining SDSC in 2001, she has been a principal investigator and a technical leader in a wide range of cross-disciplinary projects. With a specialty in scientific workflows, she leads collaborative teams to deliver impactful results through making computational data science work more reusable, programmable, scalable, and reproducible. Her work has been applied to many scientific and societal domains including bioinformatics, geoinformatics, high-energy physics, multi-scale biomedical science, smart cities, and smart manufacturing.