Q&A

David Bader
Distinguished Professor and Director of NJIT's Institute for Data Science

David Bader, distinguished professor of computer science, is the director of NJIT's new Institute for Data Science. His interests lie at the intersection of data science and high-performance computing, with applications in cybersecurity, massive-scale analytics and computational genomics. Dr. Bader works closely with researchers in academia, industry and government to develop the next generation of computing capabilities and has advised the White House on the National Strategic Computing Initiative.

One of only 32 polytechnic universities in the United States, New Jersey Institute of Technology (NJIT) prepares students to become leaders in the technology-dependent economy of the 21st century. Rated an "R1" research university by the Carnegie Classification, which indicates the highest level of research activity, NJIT conducts approximately $170 million in research activity each year. NJIT is ranked #1 nationally by Forbes for student upward economic mobility and #43 out of more than 4,000 colleges and universities by PayScale.com for the mid-career earnings of graduates. NJIT is ranked a top 100 national university by U.S. News & World Report.

What is NJIT's new Institute for Data Science?

The growing abundance and variety of data we amass gives us unprecedented opportunities to improve lives in multifold arenas — manufacturing, health care, financial management, data protection, food safety and traffic navigation are just a few. The Institute for Data Science (IDS) will focus NJIT's multidisciplinary research and workforce skills training on developing technology leaders who will solve global challenges involving data and high-performance computing (HPC). Within the Institute, collaboration among our existing research centers in big data, medical informatics and cybersecurity and our new centers in data analytics and artificial intelligence will generate data-driven technologies to achieve our goals.

How will NJIT's new Master's in Data Science advance these efforts?

We will train our master's students to think about what questions to ask of data, how to formulate analytics to answer them, to develop high-performance machine learning, and to design new techniques to turn data into real-world intelligence. By engaging confidently with complex data science tasks, our graduates will make a difference in organizations large and small. In business, for example, they will help companies compete in the global economy by harnessing a range of data in new ways: to make clear how policies affect every aspect of their enterprise, to develop transnational supply chains, and to discover efficiencies across systems.

What new capabilities will high-performance computing deliver?

We are developing predictive analytics — the use of data to anticipate the future. Instead of understanding what has happened, we wish to predict what will happen. In cybersecurity, for instance, we would create cyber analytics to defend our critical infrastructure from attack, rather than perform forensic analyses of log files after a breach. In health informatics, we want to detect diseases in their early stages and develop personalized medicines to cure them. In manufacturing, we would identify defects before they cause catastrophic failures.

How must we rethink fundamental aspects of computing to enable these capabilities?

Big data analysis is used to analyze problems related to massive datasets. Today, these datasets are loaded from storage into memory, manipulated and analyzed using HPC algorithms, and then returned in a useful format. This end-to-end workflow provides an excellent platform for forensic analysis; there is a critical need, however, for systems that support decision-making with a continuous workflow. Our HPC systems must focus on ingesting data streams; incorporating new microprocessors and custom data science accelerators that assist with loading and transforming data; and accelerating performance by moving key data science tasks and solutions from software to hardware. These workflows must be energy-efficient and easy to program, while reducing transaction times by orders of magnitude. Analysts and data scientists must be able to ask queries in their subject domain and receive rapid solutions that execute efficiently, rather than requiring sophisticated programming expertise.

Are researchers at the Institute working on these problems?

In collaboration with NVIDIA, a leading technology company that makes GPU accelerators such as the DGX Deep Learning server, we are contributing to RAPIDS.ai, an open GPU data science framework for accelerating end-to-end data science and analytics pipelines entirely on GPUs. The hardware-software co-design for analytics is exciting as we enter a new era with the convergence of data science and high-performance computing. These new analytics pipelines are more energy-efficient and run significantly faster, which is critical for making swift, data-driven decisions.

NJIT
New Jersey Institute of Technology