Georgia Tech

High Performance Computing & Data Analytics

BIG DATA

Everybody's talking about Big Data these days. The term is so hot that it's now the name of an alternative rock band. While everyone loves a good buzzword, at Georgia Tech we leverage our expertise in high performance computing and analytics to develop new methods that transform Big Data into *value*.

Every day the Internet creates a tsunami of data – usually larger than the tsunami from the day before. However, when we apply sophisticated analytics and visualization techniques to this tidal wave of information, we don't just see the tsunami: We see the currents flowing deep beneath the surface. We see "weather patterns" in the form of user behavior. We look at a turbulent, rolling ocean of ones and zeroes and discover the kinds of insights that lead to real impact on some of the most critical issues facing society today: energy and sustainability, human health and disease, and national – even international – security.

Georgia Tech's work in Big Data involves many campus units, spanning colleges, departments, and individual labs. Together, these researchers work to create new collaborative opportunities, strengthen partnerships with industry and government, and maximize the societal impact of transformative Big Data research.

All of the above is spearheaded by our School of Computational Science and Engineering (CSE), which I'm privileged to lead. Georgia Tech's School of CSE is founded on the notion that computing has joined theory and experimentation as a vital third component of modern scientific inquiry. We draw on concepts from computer science, applied mathematics, science, and engineering to create a cohesive body of knowledge to govern the representation and manipulation of models – meaning, we put HPC and data analytics to work in service of nearly all Georgia Tech's scientific and engineering disciplines.

At Georgia Tech, we're working hard to make sense of all that data. We welcome anyone – student, researcher, industry partner – who's ready to roll up their sleeves and get to work alongside us.

David Bader
Professor & Chair
School of Computational Science and Engineering
Executive Director, High Performance Computing

" ...at Georgia Tech we leverage our expertise in high performance computing and analytics to develop new methods that transform Big Data into value."

Cybersecurity
Big Data and analytics are becoming increasingly prominent in cybersecurity research and implementation. Georgia Tech researchers have developed a variety of data tools and systems to identify and assess threats, some of which have matured into commercial applications.

Data Analytics
Alternatively known as advanced analytics, predictive analytics, or machine learning, data analytics uses automated algorithms and visualization techniques to find and evaluate patterns in data, enabling increasingly accurate predictions with incredibly faster results.

Foundations
Mathematical and computing foundations that support advanced programming and the analysis of algorithms are instrumental to high-end scientific computing research. Georgia Tech's researchers investigate new methods for solving complex problems, evaluating massive sums of data, and discovering subtle patterns in data. This work provides the foundation on which all other research areas in Big Data are built.

High Performance Computing
As the demand for processing power grows, HPC systems will also need to evolve. Georgia Tech HPC research focuses on developing new parallel methods to run advanced application programs efficiently, reliably, and quickly.
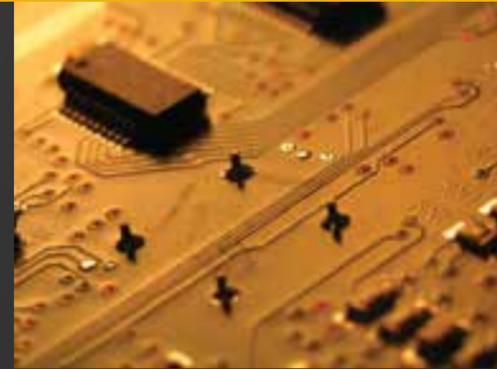
Modeling & Simulation
Modeling and simulation techniques allow researchers to simultaneously display and manage numerous data variables, bringing together information from a wide variety of sources to explore causality and interconnectedness.

Optimization
Focusing on large-scale optimization problems from a variety of industries and applications, and using mathematical models such as interior-point algorithms, Georgia Tech researchers identify new methods that enable systems to operate as efficiently and effectively as possible.
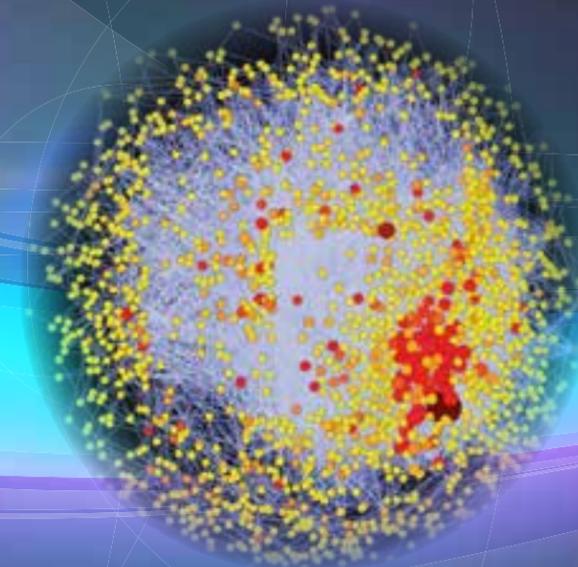
Every day in dozens of labs across the Georgia Tech campus and beyond, our researchers are tackling HPC and Big Data projects of all shapes and sizes. Here are just a few of the highlights.

XDATA: New Directions in Machine Learning for Big Data
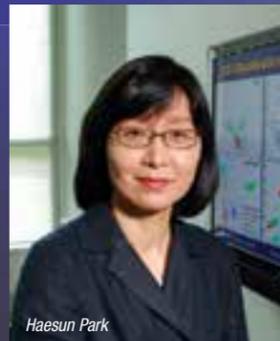
A research team led by Professor **Haesun Park** in the School of Computational Science and Engineering received a $2.7 million DARPA award to develop new computational techniques and open-source software tools for processing and analyzing data.

Haesun Park

Materials Improvement Through HPC

Advanced metal alloys have become indispensable in various emerging technologies, but Professor **Surya Kalidindi** of the Woodruff School of Mechanical Engineering knows that developing new alloys requires more than familiarity with metals chemistry. His solution is a comprehensive database of materials that can virtually compare three-dimensional microstructures to determine composite characteristics. Just one potential payoff is enhanced fuel efficiency for vehicles made lighter through the use of new alloys.

Surya Kalidindi

PERFECT-ing Power Efficiency

Under DARPA'S Power Efficiency Revolution for Embedded Computing Technologies (PERFECT), Professor and School of CSE Chair **David Bader** and Research Scientist **Jason Riedy** are creating the algorithmic infrastructure for portable supercomputing systems with a performance to rival much larger, immobile systems and requiring much less energy. The three-phase, $2.9 million project is titled GRATEFUL for Graph Analysis Tackling power-Efficiency, Uncertainty, and Locality.

Jason Riedy

XScala: Graph Analytics for Security and Social Networks

XScala, a joint project between Georgia Tech and the University of Southern California, seeks to develop a software framework for efficient accelerator kernels. Over the course of three years and funded with nearly $2 million from the NSF, researchers like Associate Professor **Richard Vuduc** and Research Scientist **Jason Riedy** will focus on diverse graph analytics for key areas such as security and social network analysis.

Big Data Analytics for DNA Sequencing

Professor **Srinivas Aluru** in the School of CSE is leading a $2 million NIH project to provide parallel bioinformatics methods for high-throughput DNA sequencing. Aluru's work addresses grand challenges in human genetics, agricultural biotechnology, and biological threat mitigation.

Srinivas Aluru

Computational Phenotyping for Smart, Connected Health

In a multi-institutional project funded with $2.1 million from the NSF, Associate Professor **Jimeng Sun** of the School of CSE will work with colleagues at Vanderbilt, Northwestern, and the University of Texas-Austin to develop algorithms and systems for automatically converting electronic health records into clinical phenotypes for research and decision making.

Jimeng Sun

Optimizing Power & Efficiency for HPC

CSE Associate Professor **Richard Vuduc** is exploring algorithms and software that are not just fast but power- and energy-efficient. Vuduc's team discovered that an "energy roofline" helps predict the best way to balance the power and performance advantages from high-demand systems with low-power alternatives for data-intensive workloads.

Richard Vuduc

Global Project Management Through GENI

Launched by the NSF in 2007, the Global Environment for Networking Innovations (GENI) initiative has a highly ambitious goal: to invent and demonstrate a global communications network that will be substantially superior to today's Internet. At Georgia Tech, Professor **Sandra Slaughter** of the Scheller College of Business focuses on GENI's project management implications, including which existing tools and techniques can be applied or tailored to fit GENI's needs.

Sandra Slaughter

System Support to Conquer Big Data

In a series of projects funded by multiple federal agencies, Professor **Karsten Schwan** of the School of Computer Science works with colleagues at Georgia Tech and around the country to examine the interplay between energy use and data movement in extreme-scale scientific computing. The "Hobbes" exascale-operating-system project compartmentalizes applications to address their specific needs, while an I/O performance project instruments the I/O and analytics workflows associated with exascale codes. Finally, the "RSVP" project develops new data management technologies and evaluates new hardware like non-volatile memory.

Karsten Schwan

BIG DATA

In recent years, tremendous advancements in Big Data technology have enabled a new generation of applications in a diversity of industries. Georgia Tech leverages its strengths and resources by reaching across traditional boundaries to embrace a multidisciplinary approach to solving today's grand challenges. We concentrate our research and tech-transfer efforts on commercial and consumer market segments that may especially benefit from Big Data research, including:

- Manufacturing & Materials
- Energy & Sustainability
- Transportation
- Security
- Bioscience & Biomedicine
- Health care

Partnership Program

At Georgia Tech, we transform world-class research into real value for our industry partners. Indeed, much of our success comes from active collaboration with partners who help drive research outcomes that are readily implemented in the commercial sector. With more than 50 current industry partners, our Big Data researchers work with companies of all sizes on short- and long-term projects, ranging from several months to a few years.

Our partners play an active role in Georgia Tech's strategy to efficiently deliver innovative concepts from the laboratory to the marketplace. Among the benefits to those partners are:

- Access to Georgia Tech laboratories for demonstrations and visits
- Meetings with faculty, students, and staff to discuss research problems
- Student-recruitment opportunities for internships, co-op positions, and permanent employment
- Access to results of all non-proprietary research

Industry and Research Foundations

Microsoft
intel
ExxonMobil
Google
NVIDIA.
LOGICBLOX®
IBM
BSF
United States – Israel Binational Science Foundation
CRAY THE SUPERCOMPUTER COMPANY
Y!
Reservoir Labs
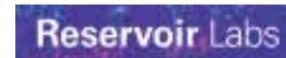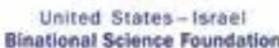
Academic excellence is Georgia Tech's cornerstone. We take a multidisciplinary, research-oriented, hands-on approach to education, where knowledge isn't the end but rather the means through which problems are solved, discoveries are made, and opportunities are created. Tech's academic programs are consistently recognized among the world's best and prepare students for success in a century where technology touches every aspect of our lives.

Degree Programs

Ph.D. in Computational Science & Engineering
The doctoral curriculum in CSE is designed to provide students with the practical skills and theoretical understanding they'll need to become leaders in the field. In conjunction with foundational courses, students develop skills to become proficient software developers on both conventional and high-performance computing platforms.

M.S. in Computational Science & Engineering
The computational science and engineering (CSE) master's degree curriculum imparts students with solid foundational knowledge and skills and includes technical specialization courses that enhance a student's domain expertise. Exposure to other disciplines teaches students how common computational science and engineering concepts are applied in different domains.

M.S. in Analytics
The new one-year Master of Science in Analytics is an interdisciplinary program that leverages Georgia Tech's nationally ranked College of Engineering, College of Computing, and Scheller College of Business. Graduates will enter the workplace with the computing, business, statistics, and operations research skills needed to identify, analyze, and solve analytics problems for better business intelligence and decision support.

Computer Science
Georgia Tech's Top 10 College of Computing offers bachelor's, master's, and Ph.D. programs in computer science that offer flexibility both in curriculum design and accessibility. Both our on-campus programs and the recently launched online MS in Computer Science (OMS CS) prepare exceptional students for careers at the cutting edge of academia, industry, and government.

Training initiatives

FLAMEL
The From Learning, Analytics, and Materials to Entrepreneurship and Leadership program—or FLAMEL—is a Ph.D. training initiative designed to accelerate the creation of new high- performance materials for applications in computing, mathematics, material science, and manufacturing. Funded by a $2.8 million NSF grant, FLAMEL will provide 24 two-year traineeships to doctoral students over five years from 2014 to 2019.

CRUISE/Data Science for Social Good
Georgia Tech's School of Computational Science and Engineering offers two 10-week summer programs that provide advanced training in computing research and application. The Computing Research Undergraduate Intern Summer Experience (CRUISE) encourages undergraduate and M.S. students to consider doctoral studies in CSE through research experiences with school faculty. The new Atlanta Data Science for Social Good (DSSG) program is an intensive, 10-week, paid internship experience that places students on multidisciplinary, faculty-supervised teams to work on problems provided by partners in the City of Atlanta and local non-profits. Mentors from the local data-science community provide additional guidance and support.

GeorgiaInstitute of Technology

Georgia Institute of Technology
High Performance Computing & Data Analytics
266 Ferst Drive NW
Atlanta, GA 30332-0280
bigdata.gatech.edu