ROCKSTARS OF HPC: David Bader

I've known David Bader for a number of years, so it was no surprise when his name came up repeatedly as a true Rock Star of HPC. We caught up with David at SC11, and he was happy to share his perspectives on supercomputing today.

"As a professor and executive director of high performance computing in the School of Computational Science and Engineering, College of Computing, at Georgia Institute of Technology, I'm honored to be selected as a Rock Star of HPC!"

insideHPC: What got you interested in high performance computing?

David Bader: I've been fascinated with computing since using my first computer, a PDP-11/45 back in 1972. As a hobbyist, I thought about solving computational challenges, and joined in on early high performance computing (HPC) discussions on early online groups such as comp.parallel several decades ago. By the time I entered high school, I had programmed on several HPC architectures and was hooked. The first parallel computer I "owned" was a 128-processor NCUBE, which a leading telecommunications company donated to my undergraduate department and left it on the doorstep with the root password taped to it. My graduate research in electrical and computer engineering focused on solving combinatorial problems using parallel computers, where I designed some of the fastest algorithms for solving large-scale discrete optimization and graph theoretic problems. HPC is an exciting technology that continues to revolutionize itself, and enables applications that could not be solved otherwise. In the early 1990's I used HPC to solve computationally-intensive problems in biology, such as assembling the most complete mouse EST database to date. At the time, HPC was thought of as a tool applicable for physics and chemistry, yet broadening the domains to include solving real-world problems in computational biology and genetics, and those in medicine and health, are what makes HPC one of the most exciting areas of research this century.

insideHPC: You have such a rich history in this community and have been involved in so many publications around HPC – what would you call out as one or two of the high points of your career – some of the things of which you are most proud?

David Bader: I am flattered to be considered a Rock Star of HPC and am privileged to have collaborated with so many amazing people, from faculty world-wide, to graduate and undergraduate students, even high school students. The highlight of my career is when my students excel in theirs. My first research experience began as a "Research Experience for Undergraduates" (REU) from the National Science Foundation, when I designed and implemented high performance algorithms to find error-correcting codes used for decoding in cryptography. The REU sparked my interest in research, and I hope to attract new researchers to our community by mentoring of the next generation.

In terms of research, one highlight is when we discovered a linear time algorithm for finding the "inversion" distance of signed permutations. Not only did we close a problem that Bill Gates opened in his only technical paper -- on the pancake flipping problem, from the mid-1970's after he dropped out of Harvard to start Microsoft in Albuquerque, NM! -- but applied this result to discover a fast and scalable way to find the evolutionary histories of genomes that undergo rearrangements. Our resulting HPC software for genome rearrangements is broadly used by pharmaceutical companies, federal agencies, Smithsonian Institute, biologists, after we demonstrated that we could find the evolutionary histories of a family of bluebell flowers in minutes using HPC rather than the previous estimates of several centuries to solve the same problem. Today we are continuing on this research, with an NSF PetaApps award to find the evolutionary histories of multi-chromosome organisms. Under an NIH award, we produced PASQUAL, our new de novo parallel genome assembly tool that is faster and more accurate than existing assemblers.

insideHPC: There was a story recently about how Big Data is being used to detect when federal employees about to go "breaking bad." Can you tell us about your work in this area?

David Bader: While HPC tends to focus on compute-intensive problems, Big Data challenges require novel architectures for data-intensive computing. My group has been the first to parallelize and implement large-scale graph theoretic algorithms, which are quite a challenge because of the irregular memory accesses, little computation to overlap with these memory references, and fine grain synchronization. In the past several years, our research has enabled social scientists to analyze some of the largest social networks, detecting communities, finding the proverbial "needle in the haystack", and "connecting the dots" by identifying central actors hidden in these networks. As you know, data and social media are now torrential streams of information that may provide valuable information to make decisions related to business intelligence, market analysis, and social trends.

When we here about stories about soldiers in good mental health who go on a shooting spree or government workers who abuse access privileges to classified data to the detriment of national security, we often wonder how these trusted individuals go unnoticed, sometimes over the period of years, and that no one detected the clues that led to these egregious incidents.

With support from the Defense Advanced Research Projects Agency (DARPA) and the Army Research Office, our team is developing new approaches for identifying "insider threats" before an incident occurs using streaming graph analytics. We are creating a suite of algorithms that can detect multiple types of insider threats by analyzing massive amounts of data for anomalous activity.

Analysts looking at the electronically recorded activities of employees within government or defense contracting organizations for anomalous behaviors may now have the bandwidth to investigate five anomalies per day out of thousands of possibilities. Our goal is to develop a system that will provide analysts for the first time a very short, ranked list of unexplained events that should be further investigated. This prototype Anomaly Detection at Multiple Scales (ADAMS) system could revolutionize the capabilities of counter-intelligence community operators to identify and prioritize potential malicious insider threats against a background of everyday cyber network activity.
This project exemplifies the hallmark of our research: solving real-world challenges. As a founding member of the School of Computational Science and Engineering at Georgia Tech, and with a joint appointment in the Georgia Tech Research Institute, our core mission is to solve problems that improve the human condition, from personalized medicine, to sustainability, and for national security and homeland defense.

insideHPC: Do you still write code?

David Bader: Once a hacker, always a hacker! The new parallel languages and systems are quite exciting, and I always have my hands dirty with novel coding, whether it is programming a new Intel manycore chip, giving IBM's X10 language a test drive, accelerating a social networking algorithm using NVIDIA's CUDA for GPUs, or solving a massive graph problem with Cray's multithreaded programming for the XMT. Some new problem always captures my interest and imagination, and I keep my skills honed coding up new parallel algorithms from applications ranging from social networks to genomics.

insideHPC: What is your favorite way to spend time when you're not working?

David Bader: Thinking about research! On a serious note, high performance computing is my passion, and I free my mind to discover new connections and solutions to real-world problems all of the time. In fact, I do not even consider my research as "work". (Please don't tell my employer, but) I cannot believe I get paid to do what I do! Often when I'm outside of the office, for instance, when I'm reading a good book or listening to music, those are the times when great moments of epiphany come to me and result in breaking through a research hurdle or finding an elegant, new way to solve a problem. When Facebook isn't drawing me in, I raise my daughter Sadie who loves to climb trees and shows me how to see the world in a different way, cook spicy vegan dishes from cuisines around the world, and secretly long to follow the Grateful Dead on tour again.

insideHPC: What motivates you? What is your passion?

David Bader: Every day when I wake up in the morning, I think about what I can do that day to help people: for example, mentoring our next-generation of students, feeding children in impoverished nations, or keeping populations safe from the spread of disease or the potential of war. The most interesting research areas are at the interface between disciplines, and I always want the uncomfortable feeling that one gets when focusing on these emerging frontiers. I am passionate about solving the real problem at hand, not some computer science abstraction or small unit that will get the next academic paper out the door, but truly making an impact in the world that will be lasting and profound. One should never focus on the traditional measures of success in academia – number of papers published, citation counts, dollars of funded awards, etc. – rather, we should follow our passions, and success is truly knowing that we made a difference in the world.

insideHPC: What do you see as the most exciting possibility of what we can hope to accomplish over the next 5-10 years through the innovative applications of HPC?

David Bader: I'm optimistic that HPC will lead to innovations in personalized medicine and health care by providing a better understanding of the genome and life's processes, will augment our social understanding of each other and cultures around the world, and will improve national security so that our children inherit a safer world.