NATIONAL SCIENCE FOUNDATION (NSF)
RESEARCH EXPERIENCES FOR UNDERGRADUATES (REU) SITE

EXERCISE - Explore Emerging Computing in Science and Engineering
Research Speaker Series

High-Throughput Data-Intensive Computing:
I/O Sharing in Scientific Databases and the Cloud

Professor Randal Burns

Date: Thursday June 7, 2012

Time: 10:30am-11:30am

Location: Henson Science 101

Abstract: Data-intensive computing consists of batch-processing workloads that scan massive data sets in parallel. The focus on data access, data movement, data ingest, and data production means that these workloads overwhelm the network and I/O capabilities of data centers and supercomputers. Major improvements in throughput are available by co-scheduling tasks that access the same data so that multiple tasks complete processing based on accessing and transferring the data a single time. Multiple tasks share I/O, network data transfer, cache space, and even computing. This talk will review the evolution of co-scheduling in data-intensive computing systems, including shared-scan scheduling for map/reduce workloads (Agrawal et al., VLDB 2008), data-driven batch processing for scientific databases (LifeRaft and JAWS), shared streaming-I/O for spatial workloads, and shared join processing for Pig programs and Nova workflows.

Bio: Randal Burns is an Associate Professor of Computer Science at the Johns Hopkins University and a founding member of the Johns Hopkins Institute for Data-Intensive Science and Engineering (IDIES). His research interests include storage and memory systems, high-performance computing, and scientific computing. But mostly, he helps scientists deal with uncomfortably large data sets; his group operates several Open-Science Web-services including the Open Connectome Project (openconnecto.me) and the JHU Turbulence Database Cluster (turbulence.pha.jhu.edu). Randal was formerly a Research Staff Member in Storage Systems at IBM's Almaden Research Center in San Jose. Randal earned his Ph.D. in 2000 and M.S. in 1997 from the Department of Computer Science at the University of California at Santa Cruz. He earned his B.S. degree from the Department of Geophysicsat Stanford University.

For more information, visit online: http://faculty.salisbury.edu/~ealu/REU/REU.html
or contact Dr. Enyue (Annie) Lu: Email: ealu@salisbury.edu Phone: (410) 543-6144