

The 2nd International Workshop on  
**Network-aware Data Management**

to be held in conjunction with the IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis (SC12)



*Sunday, November 11th, 2012 (9:00am - 5:30pm)*

Room 155-A

Salt Palace Convention Center, Salt Lake City, Utah

Scope: Scientific applications and experimental facilities generate large amounts of data. In addition to increasing data volumes and computational requirements, today's major science requires cooperative work in globally distributed multidisciplinary teams. In the age of extraordinary advances in communication technologies, there is a need for efficient use of the network infrastructure to address increasing data and compute requirements of large-scale applications. Since the amount of data and the size of scientific projects are continuously growing, traditional data management techniques are unlikely to support future collaboration systems at the extreme scale. Network-aware data management services for dynamic resource provisioning, end-to-end processing of data, intelligent data-flow and resource coordination are highly desirable. This workshop seeks contribution from academia, government, and industry to discuss emerging trends in use of networking for data management, novel techniques for data representation, simplification of end-to-end data flow, resource coordination, and network-aware tools for the scientific applications.

Keynote Speech: **Data-intensive and Cloud Applications in Large-scale Data Center Systems**

Karsten Schwan, Georgia Institute of Technology

*Abstract:* Data-intensive applications have been evolving from their original focus on offline mining of business data into broader domains, including the online inspection and analysis of large-scale web data used for rapid response to current conditions. Processing such 'data in motion' brings new challenges to the domain of data intensive computing. This talk will articulate some of those challenges, present representative solutions, and describe potential avenues for future work, in lieu of several constraints seen for this broad class of datacenter applications, including their use of shared underlying datacenter infrastructure, their support by datacenter operators, and the time-constrained operation inherent in their execution. Future research opportunities in this space include application acceleration via GPGPUs as well as new ways to enrich the open source infrastructures used to run these codes.

Invited Talk: **Optimizing Transport of Big Data over Dedicated Networks**

Dipak Ghosal, University of California, Davis

*Abstract:* Data centers are being deployed in cloud computing environments, scientific, financial, defense, and other enterprises. Geographically distributed data centers transmit and receive growing volumes of data. In order to avoid congestion in the public Internet, they use high speed dedicated optical networks, which can be thought of as private highways for carrying data. A careful examination of the impact of such high speed network traffic on a commodity multicore machine show packet loss and degraded throughput due to the end-system being the bottleneck. We found that high-speed single flow traffic nullifies the benefits of multicore systems and multiqueue NICs. We propose an end-system aware flow control technique to optimize the data transfer time using rate-based protocols. Using introspective end-system modeling, we show that we can determine the optimal number of parallel flows required to utilize the available bandwidth and the optimal rate for each of the flows.

Panel Discussion: **New Directions in Networking and Data Management**

Panelists:

Ali R. Butt, Virginia Tech  
Zihui Du, Tsinghua University, Beijing  
Shantenu Jha, Rutgers University  
Raj Kettimuthu, Argonne National Laboratory  
Inder Monga, Energy Sciences Network  
Jason Zurawski, Internet2

Workshop Organizers: Mehmet Balman and Surendra Byna, Lawrence Berkeley National Laboratory

The 2nd International Workshop on

# Network-aware Data Management (NDM)

to be held in conjunction with the IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis (SC12)

Sunday, November 11th, 2012 (9:00am - 5:30pm)

Room 155-A



## WORKSHOP PROGRAM

09:00	Opening Remarks
<a href="#">Keynote Speech:</a>	
09:10	<b>Data-intensive and Cloud Applications in Large-scale Data Center Systems</b> Karsten Schwan, Georgia Institute of Technology
10:10	Break
<a href="#">Paper Session I:</a>	
10:30	<b>How GridFTP pipelining, parallelism and concurrency work: A guide for optimizing large dataset transfers</b> Esma Yildirim, Fatih University, Turkey (presenter); Jangyoung Kim, University at Buffalo; Tevfik Kosar, University at Buffalo
11:00	<b>Accelerating Data Movement Leveraging Endsystem and Network Parallelism</b> Jun Yi, Argonne National Laboratory; Rajkumar Kettimuthu, Argonne National Lab and The University of Chicago (presenter); Venkat Vishwanath, Argonne National Laboratory
	Invited Talk:
11:30	<b>Optimizing Transport of Big Data over Dedicated Networks</b> Dipak Ghosal, University of California, Davis
12:00	Lunch Break
<a href="#">Paper Session II:</a>	
13:00	<b>A Dynamic Virtual Networks Solution for Cloud and Grid Computing</b> Davide Salomoni, INFN CNAF, Italy; Marco Caberletti, Italy, INFN CNAF, Italy
13:20	<b>Hadoop acceleration in an OpenFlow-based cluster</b> Sandhya Narayan, InfoBlox (presenter); Stuart Bailey, InfoBlox; Anand Daga, University of Houston
<a href="#">Contributed Talk:</a>	
13:40	Streaming Exa Scale data over 100Gbps Networks, Mehmet Balman
13:55	Break
<a href="#">Paper Session III:</a>	
14:10	<b>Adaptive Data Transfers that Utilize Policies for Resource Sharing</b> Junmin Gu, Lawrence Berkeley National Laboratory; David Smith, University of Southern California Information Sciences Institute; Ann L. Chervenak, University of Southern California Information Sciences (presenter); Alex Sim, Lawrence Berkeley National Laboratory
14:30	<b>A Network-aware Object Storage Service</b> Shigetoshi Yokoyama, National Institute of Informatics, Japan (presenter); Nobukazu Yoshioka, National Institute of Informatics, Japan; Motonobu Ichimura, NTT DATA Intellilink, Japan
14:50	<b>Efficient Attribute-based Data Access in Astronomy Analysis</b> Benson Ma, Lawrence Berkeley National Laboratory (presenter); Arie Shoshani, Lawrence Berkeley National Laboratory; Alex Sim, Lawrence Berkeley National Laboratory; Kesheng Wu, Lawrence Berkeley National Laboratory; Yong-Ik Byun, Yonsei University, Korea; Jaegyeon Hahm, Institute of Science and Technology Information, Korea; Min-Su Shin, University of Michigan
15:10	Break
15:30	<b>Panel Discussion (New Direction in Networking and Data Management)</b> Moderators: Mehmet Balman and Surendra Byna  Panelists: Ali R. Butt, Virginia Tech; Zhihui Du, Tsinghua University, Beijing; Shantenu Jha, Rutgers University; Raj Kettimuthu, Argonne National Laboratory; Inder Monga, Energy Sciences Network; Jason Zurawski, Internet2
17:30	Closing Remarks & Best paper award

Workshop Organizers: Mehmet Balman and Surendra Byna, Lawrence Berkeley National Laboratory