

Previously Held: NDM 2012

The Second International Workshop on Network-aware Data Management held on Nov 12th, 2012, in conjunction with the International Conference for High Performance Computing, Networking, Storage and Analysis (SC 2012), in Salt Lake City, Utah, USA.

The goal of the one-day workshop is to seek contributions from academia, government, and industry to discuss emerging trends in network-aware data management. Topics of interest included end-to-end processing of data, novel techniques for data representation, simplification of end-to-end data flow, resource coordination, high-performance data movement, and network-aware tools for the scientific applications. The call for papers received submissions from Asia, Europe, and North America.

The format of the workshop consisted of a keynote speech, an invited talk, presentations of peer-reviewed papers, a contributed talk, and a panel discussion. The review process involved at least 3 reviews for each submission. We ranked the review scores and selected high-quality papers that best fit into the workshop program. Based on the review results, we accepted 8 papers.

One Best-Paper was selected as an outstanding paper contribution. The paper titled “*How GridFTP pipelining, parallelism and concurrency work: A guide for optimizing large dataset transfers*” co-authored by Esma Yildirim (Fatih University, Turkey), Jangyoung Kim and Tefvik Kosar (University at Buffalo SUNY, USA) received the Best Paper Award. We announced the awards during the closing remarks, and presented certificates to the authors at the end of the workshop program.

The program includes a panel on “*New Directions in Networking and Data Management*”, a keynote speech titled “*Data-intensive and Cloud Applications in Large-scale Data Center Systems*” by Karsten Schwan (Georgia Institute of Technology), and an invited talk titled “*Optimizing Transport of Big Data over Dedicated Networks*” by Dipak Ghosal (University of California, Davis). We invited 6 panelists from different specialties who are well-known in the field of networking and data-intensive computing. Panel members include Ali R. Butt (Virginia Tech), Zhihui Du (Tsinghua University, Beijing), Shantenu Jha (The State University of New Jersey, Rutgers), Rajkumar Kettimuthu (Argonne National Laboratory), Inder Monga (Energy Sciences Network), and Jason Zurawski (Internet2).

We would like to thank the NDM 2012 authors for providing high-quality paper contributions. We also thank our program committee members for reviewing papers and providing feedback to the authors. Finally, we thank all the panelists and speakers for their insightful comments and presentations.

We posted the presentation slides from paper presentations, panel talks, invited talk, and keynote speech online at <http://sdm.lbl.gov/ndm/2012>

Workshop Website: <http://sdm.lbl.gov/ndm/2012>

NDM 2012 Program

09:00 Opening Remarks

Keynote Speech:

09:10 **Data-intensive and Cloud Applications in Large-scale Data Center Systems**
Karsten Schwan, Georgia Institute of Technology

10:10 Break

Paper Session I:

10:30 **How GridFTP pipelining, parallelism and concurrency work: A guide for optimizing large dataset transfers**

Esma Yildirim, Fatih University, Turkey (presenter); Jangyoung Kim, University at Buffalo; Tevfik Kosar, University at Buffalo

11:00 **Accelerating Data Movement Leveraging Endsystem and Network Parallelism**

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 **Optimizing Transport of Big Data over Dedicated Networks**
Dipak Ghosal, University of California, Davis

12:00 Lunch Break

Paper Session II:

13:00 **A Dynamic Virtual Networks Solution for Cloud and Grid Computing**

Davide Salomoni, INFN CNAF, Italy; Marco Caberletti, Italy, INFN CNAF, Italy

13:20 **Hadoop acceleration in an OpenFlow-based cluster**

Sandhya Narayan, InfoBlox (presenter); Stuart Bailey, InfoBlox; Anand Daga, University of Houston

Contributed Talk:

13:40 Streaming Exa Scale data over 100Gbps Networks, Mehmet Balman

13:55 Break

Paper Session III:

14:10 **Adaptive Data Transfers that Utilize Policies for Resource Sharing**

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 **A Network-aware Object Storage Service**

Shigetoshi Yokoyama, National Institute of Informatics, Japan (presenter); Nobukazu Yoshioka, National Institute of Informatics, Japan; Motonobu Ichimura, NTT DATA Intellilink, Japan

14:50 **Efficient Attribute-based Data Access in Astronomy Analysis**

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; Jaegyoon Hahm, Institute of Science and Technology Information, Korea; Min-Su Shin, University of Michigan

15:10 Break

Panel Discussion (**New Direction in Networking and Data Management**)

15:30 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

The 2nd International Workshop on
Network-aware Data Management

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

Zhihui Du, Tsinghua University, Beijing

Shantenu Jha, Rutgers University

Raj Kettimuthu, Argonne National Laboratory

Inder Monga, Energy Sciences Network

Jason Zurawski, Internet2

NDM 2012 Panel Discussion

Topic: **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

Suggested Panel Questions:

- What do you envision will have dramatic impact in the future networking and data management? What research challenges do you expect in achieving your vision?
- Do we need to re-engineer existing tools and middleware software? Elaborate on network management middleware in terms of virtual circuits, performance monitoring, and diagnosis tools.
- How do current applications match increasing data sizes and enhancements in network infrastructure? Please list a few network-aware application. What is the scope of networking in the application domain?
- Resource management and scheduling problems are gaining importance due to current developments in utility computing and high interest in Cloud infrastructure. Explain your vision. What sort of algorithms/mechanisms will practically be used in the future?
- What are the main issues in designing/modelling cutting edge dynamic networks for large-scale data processing? What sort of performance problems do you expect?
- What necessary step do we need to implement to benefit from next generation high bandwidth networks? Do you think there will be radical changes such as novel APIs or new network stacks?

NDM 2012 Organization

General Chairs:

Mehmet Balman (Lawrence Berkeley National Laboratory, USA)

Surendra Byna (Lawrence Berkeley National Laboratory, USA)

Program Committee:

Ismail Akturk (Bilkent University, Turkey)

Gabrielle Allen (Louisiana State University, USA)

Pavan Balaji (Argonne National Laboratory, USA)

Ali R. Butt (Virginia Tech., USA)

Tasneem Brutch (Samsung R&D, USA)

Promita Chakraborty (Molecular Foundry, LBNL, USA)

Constantine Dovrolis (Georgia Tech., USA)

Zhihui Du (Tsinghua University, China)

Alexandru Iosup (Delft University of Technology, the Netherlands)

Daniel S. Katz (University of Chicago, USA)

Dhabaleswar Panda (Ohio State University, USA)

Manish Parashar (The State University of New Jersey, Rutgers, USA)

Ian Taylor (Cardiff University, United Kingdom)

Sudharshan Vazhkudai (Oak Ridge National Laboratory, USA)

Venkatram Vishwanath (Argonne National Laboratory, USA)

Jason Zurawski (Internet2, USA)

Fatos Xhafa (University of Catalonia, Spain)