HPDC Workshop Programs

Workshops on Monday, June 21st:	Time	Location
Workshop on Scientific Cloud Computing	8:45 - 6:15	Superior 1
Emerging Computational Methods for the Life Sciences	8:30 - 6:00	Superior 3
Managing Data Quality in Collaborative Science	8:30 - 12:30	Superior 2
Large-Scale System and Application Performance	1:30 - 5:30	Superior 2

Workshops on Tuesday, June 22nd:	Time	Location
Workshop on MapReduce and Applications	9:00 - 5:00	Superior 1
Virtualization Technologies for Distributed Computing	9:00 - 5:00	Superior 2
Data Intensive Distributed Computing	9:00 - 5:00	Superior 3
Challenges of Large Applications in Distributed Environments	9:30 - 5:00	Michigan

Workshop on Scientific Cloud Computing
Monday, June 21st in Room Superior 1

Workshop Chairs: Peter Beckman, Ian Foster, Ioan Raicu

8:45   Welcoming Remarks

8:50   Keynote Talk: The Client + Cloud: Changing the Paradigm for Scholarly Research
Dennis Gannon, Microsoft Research

9:30   Early Observations on the Performance of Windows Azure
Arkaitz Ruiz-Alvarez, Zach Hill, Ming Mao, Jie Li, and Marty Humphrey

10:00   An Overview of the Open Science Data Cloud
Robert Grossman, Yunhong Gu, Michal Sabala, Joe Mambretti, Alex Szalay, and Kevin White

10:30   Coffee Break

11:00   Providing a Cloud Network Infrastructure on a Supercomputer
Eric Van Hensbergen, Robert Wisniewski, Bryan Rosenburg, Amos Waterland, Dilma Da Silva, Jonathan Appavoo, Volkmar Uhlig, Jan Stoess, and Udo Steinberg

11:30   Case Study for Running HPC Applications in Public Clouds
Qiming He, Shujia Zhou, Ben Kobler, Dan Duffy, and Tom McGlynn

12:00   Exploring Application and Infrastrucre Adaptation on Hybrid Grid-Cloud Infrastructure
Hyunjoo Kim, Manish Parashar, Yaakoub El Khamra, and Shantenu Jha

12:30   Lunch

1:30   AzureBlast: A Case Study of Cloud Computing for Science Applications
Wei Lu, Jared Jackson, and Robert Barga

2:00   Seeking Supernovae in the Clouds: A Performance Study
Keith Jackson, Lavanya Ramakrishnan, Rollin Thomas, and Karl Runge

2:30   Migrating a Large Science Database to the Cloud
Ani Thakar and Alex Szalay

3:00   Reshaping Text Data for Efficient Processing on Amazon EC2
Gabriela Turcu, Svetlozar Nestorov, and Ian Foster

3:30   Coffee Break

4:00   Comparison of Resource Platform Selection Approaches for Scientific Workflows
Yogesh Simmhan, and Lavanya Ramakrishnan

4:30   A Perspective on Scientific Cloud Computing
Craig Lee

5:00   Panel Discussion: Scientific Cloud Computing: Reality or Vaporware?
Manish Parashar, Dennis Gannon, Kate Keahey, Peter Dinda, Bob Grossman

6:00   Best Paper Award
6:10   Closing Remarks

Workshop: Emerging Computational Methods for the Life Sciences
Monday, June 21st in Room Superior 3

Workshop Chairs: Ian Foster, Geoffrey Fox, Judy Qiu

8:30   Microsoft Biological Framework and its Applications
Simon Mercer

9:00   Cloud Computing Paradigms for Pleasingly Parallel Biomedical Applications
Thilina Gunarathne, Tak-Lon Wu, Judy Qiu, and Geoffrey Fox

9:30   MPIPairwiseStatSig: Parallel Pairwise Statistical Significance Estimation of Local Sequence Alignment
Ankit Agrawal, Sanchit Misra, Daniel Honbo, and Alok Choudhary

10:00   Exploring the RNA Folding Energy Landscape Using Scalable Distributed Cyberinfrastructure
Joohyun Kim, Wei Huang, Sharath Maddineni, Fareed Aboul-ela and Shantenu Jha

10:30   Coffee Break

11:00   Keynote Talk:
Genomics and High Performance Computing
Folker Meyer, Argonne Mational Lab

12:00   Biocompute: Towards a Collaborative Workspace for Data Intensive Bio-Science
Rory Carmichael, Patrick Braga-Henebry, Douglas Thain, and Scott Emrich

12:30   Lunch

1:30   Modeling sequence and function similarity between proteins for protein functional annotation
Roger Higdon, Brenton Louie, and Eugene Kolker

2:00   Browsing Large Scale Cheminformatics Data with Dimension Reduction
Jong Youl Choi, Seung-Hee Bae, Judy Qiu, Geoffrey Fox, Bin Chen, and David Wild

2:30   Data Parallelism in Bioinformatics Workflows Using Hydra
Fábio Coutinho, Eduardo Ogasawara, Daniel de Oliveira, Vanessa Braganholo, Alexandre Lima, Alberto Dávila, and Marta Mattoso

3:00   Optimization of a parallel permutation testing function for the SPRINT R package
Savvas Petrou, Terence Sloan, Muriel Mewissen, Thorsten Forster, Michal Piotrowski, and Bartosz Dobrzelecki

3:30   Coffee Break

4:00   Panel Discussion: Data Intensive Cyberinfrastructure for Life Sciences

5:00   Text Mining for Bone Biology
Andrew Hoblitzell, Snehasis Mukhopadhyay, Qian You, Shiaofen Fang, Yuni Xia, and Joseph Bidwell

5:30   CUDA-based Triangulations of Convolution Molecular Surfaces
Sergio Dias and Kuldeep Bora

Workshop on Managing Data Quality for Collaborative Science
Monday, June 21st in Room Superior 2

Workshop Chairs: Todd Halter, Maria Indrawan, Eric Stephan

8:30   Keynote Talk:
Yang Lee, Northeastern University
Editor-in-Chief, Journal of Data and Information Quality

9:30   Optimal Enterprise Data Architecture Using Publish and Subscribe
Carlo Batani, Simone Grega, and Andrea Maurino

10:00   The Development of QC Standards for ARM Data Products
Krista Gaustad, Connor Flynn, Sherman Beus, and Brian Ermold

10:30   Coffee Break

11:00   A Quality Screening Service for Remote Sensing Data
Christopher Lynnes, Edward Olsen, Peter Fox, Bruce Vollmer, Robert Wolfe,
and Shahin Samadi

11:30   Monitoring Data Quality in Kepler
Aisa Naim, Daniel Crawl, Maria Indrawan, Ilkay Altintas, and Shulei Sun

12:00   Towards Long Term Data Quality in a Large Scale Biometrics Experiment
Hoang Bui, Diane Wright, Clarence Helm, Rachel Witty, Patrick Flynn, and Douglas Thain

Workshop on Large-Scale System and Application Performance
Monday, June 21st in Room Superior 2

Workshop Chairs: Dick Epema, Jose Moreira, Carey Williamson

1:30   Keynote Talk:
Scaling up to Large (Really Large) Systems
Barton P. Miller, University of Wisconsin

2:30   Netlag: A Performance Evaluation Tool for Massively Multi-User Networked Applications
Alexander Ploss, Dominik Meilander, Philipp Mollers, Frank Glinka, and Sergei Gorlatch

3:00   BTWorld: Towards Observing the Global BitTorrent File-Sharing Network
Maciej Wojciechowski, Mihai Capota, Johan Pouwelse, and Alexandru Iosup

3:30   Coffee Break

4:00   A Hybrid Markov Chain Model for Workload on Parallel Computers
Anne Krampe, Joachim Lepping, and Wiebke Sieben

4:30   LogGOPSim - Simulating Large-Scale Applications in the LogGOPS Model
Torsten Hoefler, Timo Schneider, and Andrew Lumsdaine

5:00   Fast and Scalable Simulation of Volunteer Computing Systems Using SimGrid
Bruno Donassolo, Henri Casanova, Arnaud Legrand, and Pedro Velho

Workshop on MapReduce and its Applications
Tuesday, June 22nd in Room Superior 1

Workshop Chairs: Gilles Fedak, Geoffrey Fox, Haiwu He

9:00   Parallel Processing of Data from Very Large-Scale Wireless Sensor Networks
Christine Jardak, Janne Riihijarvi, Frank Oldewurtel, and Petri Mahonen

9:30   Massive Semantic Web Data Compression with MapReduce
Jacopo Urbani, Jason Maassen, and Henri Bal

10:00   Very Large Pattern Databases for Heuristic Search
Alexander Reinefeld, Thorsten Schuett, and Robert Maier

10:30   Coffee Break

11:00   Twister: A Runtime for Iterative MapReduce
Jaliya Ekanayake, Hui Li, Bingjing Zhang, Thilina Gunarathne, Seung-Hee Bae, Judy Qiu, and Geoffrey Fox

11:30   Pydoop: A Python MapReduce and HDFS API for Hadoop
Simone Leo and Gianluigi Zanetti

12:00   Pairwise Element Computation with MapReduce
Tim Kiefer, Peter Benjamin Volk, and Wolfgang Lehner

12:30   Lunch

1:30   Keynote Talk:
MapReduce Inside Google: Implementation, Applications, and Alternatives
Jerry Zhao, Google Inc.

2:30   Improving the Hadoop Map/Reduce Framework to Support Concurrent Appends through the BlobSeer BLOB Management System
Diana Moise, Gabriel Antoniu, and Luc Bouge

3:00   Multi-GPU Volume Rendering Using Map-Reduce
Jeff Stuart, Cheng-Kai Chen, Kwan-Liu Ma, and John Owens

3:30   Coffee Break

4:00   MR-Scope: A Real-Time Tracing Tool for MapReduce
Dachuan Huang, Xuanhua Shi, Shadi Ibrahim, Lu Lu, Hongzhang Liu, Song Wu, and Hai Jin

4:30   Parallelizing Multiple Group-by Query in Share-Nothing Environment: A MapReduce Study Case
Jie Pan, Yann le Biannic, and Frederic Magoules

Workshop on Virtualization Technologies for Distributed Computing
Tuesday, June 22nd in Room Superior 2

Workshop Chairs: Renato Figueiredo, Frederic Desprez

9:00   Invited Talk:
        Virtualization Technologies in Distributed Architecture: The Grid5000 Recipe
        Adrien Lebre

10:00   Cluster-Wide Context Switch of Virtualized Jobs
        Fabien Hermenier, Adrien Lebre, and Jean-Marc Menaud

10:30   Coffee Break

11:00   Scaling Virtual Organization Clusters over a Wide Area Network
        using the Kestrel Workload Management System
        Lance Stout, Michael Fenn, Micahel Murphy, and Sebastien Goasguen

11:30   Pools of Virtual Boxes: Building Campus Grids with Virtual Machines
        David Herzfeld, Lars Olson, and Craig Struble

12:00   Janus: A Cross-Layer Soft Real-Time Architecture for Virtualization
        Raoul Rivas, Ahsan Arefin, and Klara Nahrstedt

12:30   Lunch

1:30   Invited Talk:
        An Introduction to the V3VEE Project and the Palacios Virtual Machine Monitor
        Peter Dinda

2:30   DistriBit: A Distributed Dynamic Binary Translator System for Thin Client Computing
        Haibing Guan, Yindong Yang, Kai Chen, Yindong Ge, Liang Liu, and Ying Chen

3:00   Storage Deduplication for Virtual Ad Hoc Network Testbed By File-Level Block Sharing
        *Chang-Han Jong, Cho-Yu Lason Chiang, Taichuan Lu, Alexander Poylisher,
        and Constantin Serban*

3:30   Coffee Break

4:00   Invited Talk:
        Future Grid: Supporting Next Generation Data Intensive Cyberinfrastructure
        Geoffrey Fox

Workshop on Data Intensive Distributed Computing
Tuesday, June 22nd in Room Superior 3

Workshop Chair: Tevfik Kosar

9:00   Opening Remarks
       Tevfik Kosar

9:10   Keynote Talk:
       It's not a Data Deluge – It's Worse than That
       Craig A. Stewart

10:00 Characterizing a Grid Site's Traffic
       Tiejun Ma, Yehia El-khatib, Michael Mackay, and Christopher Edwards

10:30 Coffee Break

11:00 A Data Transfer Framework for Large-Scale Science Experiments
       Wantao Liu, Brian Tieman, Rajkumar Kettimuthu, and Ian Foster

11:30 Towards Optimising Distributed Data Streaming Graphs Using Parallel Streams
       C.S. Liew, M.P. Atkinson, J.I. van Hemert, and L. Han

12:00 Detouring and Replication for Fast and Reliable Internet-Scale Stream Processing
       Christopher McConnell, Fan Ping, and Jeong-Hyon Hwang

12:30 Lunch

1:30   File-Access Patterns of Data-Intensive Workflow Applications and their Implications
       to Distributed Filesystems
       Takeshi Shibata, SungJun Choi, and Kenjiro Taura

2:00   Versioning for Workflow Evolution
       Eran Chinthaka Withana, Beth Plale, Roger Bargan, and Nelson Araujo

2:30   ROARS: A Scalable Repository for Data Intensive Scientific Computing
       Hoang Bui, Peter Bui, Patrick Flynn, and Douglas Thain

3:00   GatorShare: A File System Framework for High-Throughput Data Management
       Jiangyan Xu and Renato Figueiredo

3:30   Coffee Break

4:00   Panel Discussion: "Bringing Communities Together for Data Intensive Scientific Discovery"

5:00   Closing Remarks
       Tevfik Kosar

Challenges of Large Applications in Distributed Environments
Tuesday, June 22nd in Room Michigan

Workshop Chairs: Daniel S. Katz, Shantenu Jha

9:30   Keynote Talk:
The Earth System Grid Federation: A Globally Distributed Environment for Climate Research
Don Middleton, National Center for Atmospheric Research

10:30   Coffee Break

11:00   Efficient Querying of Distributed Provenance Stores
Ashish Gehani, Minyoung Kim, and Tanu Malik

11:30   SAGA-based File Access Application over Multi-Filesystem Middleware
Yutaka Kawai, Go Iwai, Takashi Saski, and Yoshiyuki Watase

12:00   Toward High Performance Computing in Unconventional Computing Environments
Brent Rood, Nathan Gnanasambandam, Michael Lewis, and Naveen Sharma

12:30   Lunch

1:30   Invited Talk:
Developing Science Applications on the Cloud: Case Study and Lessons Learned
Wei Lu, Microsoft

2:00   Weaver: Integrating Distributed Computing Abstractions into Scientific Workflows Using Python
Peter Bui, Li Yu, and Douglas Thain

2:30   A Distributed Workflow for an Astrophysical Open MP Application
Robert Henschel, Scott Michael, and Stephen Simms

3:00   Design, Implementation, and Use of Simulation Data Archive for Coastal Science
Harsha Bhagawaty, Lei Jiang, Sreekanth Pothanis, Gabrielle Allen, Nathan Brener, Tevik Kosar, Swathi Dubbaka, Kelin Hu, and Qin Chen

3:30   Coffee Break

4:00   Panel Discussion