# Table of Contents

Message from the General Co-Chairs ........................................................................................................... xv
Message from Technical Program Committee Co-Chairs ............................................................................ xvii
Executive Committee ........................................................................................................................................ xix
Technical Program Committee ..................................................................................................................... xxix
Reviewers ........................................................................................................................................................ xxvii
Keynotes .......................................................................................................................................................... xxix

---

## Session 1: Best Papers

### Improving I/O Throughput of Scientific Applications Using Transparent Parallel Compression
Tekin Bicer, Jian Yin, and Gagan Agrawal

### Performance Models for CPU-GPU Data Transfers
B. van Werkhoven, J. Maassen, F.J. Seinstra, and H.E. Bal

### A User-Level InfiniBand-Based File System and Checkpoint Strategy for Burst Buffers
Kento Sato, Kathryn Mohror, Adam Moody, Todd Gamblin, Bronis R. de Supinski, Naoya Maruyama, and Satoshi Matsuoka

---

## Session 2A: MapReduce

### Tagged-MapReduce: A General Framework for Secure Computing with Mixed-Sensitivity Data on Hybrid Clouds
Chunwang Zhang, Ee-Chien Chang, and Roland H.C. Yap

### Toward Detecting Compromised MapReduce Workers through Log Analysis
Eunjung Yoon and Anna Squicciarini

### MapReduce Analysis for Cloud-Archived Data
Balaji Palanisamy, Aameek Singh, Nagapramod Mandagere, Gabriel Alatorre, and Ling Liu

### Adaptive MapReduce Scheduling in Shared Environments
Jordà Polo, Yolanda Becerra, David Carrera, Jordi Torres, Eduard Ayguadé, and Malgorzata Steinder
Session 2B: Energy and the Environment

Enabling Efficient Power Provisioning for Enterprise Applications .......................................................... 71
  Balaji Subramaniam and Wu-Chun Feng

Analytical/ML Mixed Approach for Concurrency Regulation in Software Transactional Memory .......................................................... 81
  Diego Rughetti, Pierangelo Di Sanzo, Bruno Ciciani, and Francesco Quaglia

Bridging Data in the Clouds: An Environment-Aware System for Geographically Distributed Data Transfers .......................................................... 92
  Radu Tudoran, Alexandru Costan, Rui Wang, Luc Bougé, and Gabriel Antoniu

Cost-Efficient, Reliable, Utility-Based Session Management in the Cloud .................................................. 102
  Benjamin Byholm and Iván Porres

Session 2C: Resource Management

A Workflow-Inspired, Modular and Robust Approach to Experiments in Distributed Systems .......................................................... 112
  Tomasz Buchert, Lucas Nussbaum, and Jens Gustedt

Resource Usage Control in Multi-tenant Applications .................................................................................. 122
  Rouven Krebs, Simon Spinner, Nadia Ahmed, and Samuel Kounov

Mapping Algorithms Optimizing the Overall Manhattan Distance for Pre-Occupied Cluster Computers in SLA-Based Grid Environments .......................................................... 132
  Barry Linnert, Joerg Schneider, and Lars-Olof Burchard

SLA-Based Profit Optimization in Cloud Bursting PaaS ................................................................................. 141
  Djawida Dib, Nikos Parlavantzas, and Christine Morin

Session 3A: Apps I

Evaluating Streaming Strategies for Event Processing Across Infrastructure Clouds .................................................. 151
  Radu Tudoran, Kate Keahey, Pierre Riteau, Sergey Panitkin, and Gabriel Antoniu

CUDAlign 3.0: Parallel Biological Sequence Comparison in Large GPU Clusters .................................................. 160
  Edans F. de O. Sandes, Guillermo Miranda, Alba C.M.A. de Melo, Xavier Martorell, and Eduard Ayguadé

Opportunistic High Energy Physics Computing in User Space with Parrot .................................................. 170

Session 3B: Message Passing

Towards an MPI-Like Framework for the Azure Cloud Platform .......................................................... 176
  Dinesh Agarwal, Sara Karamati, Satish Puri, and Sushil K. Prasad

NoCMMsg: Scalable NoC-Based Message Passing .................................................................................. 186
  Christopher Zimmer and Frank Mueller

Modeling and Optimizing Large-Scale Wide-Area Data Transfers .......................................................... 196
  Rajkumar Kettimuthu, Gayane Vardoyan, Gagan Agrawal, and P. Sadayappan
Session 3C: Elasticity and Adaptation

A Language Support for Cloud Elasticity Management ................................................................. 206
Yousri Kouki, Frederico Alvares de Oliveira Jr., Simon Dupont, and Thomas Ledoux

Elastic MapReduce Execution ........................................................................................................ 216
Wei Xiang Goh and Kian-Lee Tan

JCatascopia: Monitoring Elastically Adaptive Applications in the Cloud ........................................... 226
Demetris Trihinas, George Pallis, and Marios D. Dikaiakos

Session 4A: Big Data

Towards a Collective Layer in the Big Data Stack ......................................................................... 236
Thilina Gunarathne, Judy Qiu, and Dennis Gannon

Flexpath: Type-Based Publish/Subscribe System for Large-Scale Science Analytics .................. 246
Jai Dayal, Drew Bratcher, Greg Eisenhauer, Karsten Schwan, Matthew Wolf, Xuechen Zhang, Hasan Abbasi, Scott Klasky, and Norbert Podhorszki

Transparent in Situ Data Transformations in ADIOS ..................................................................... 256
David A. Boyuka, Sriram Lakshminarasimham, Xiaocheng Zou, Zhenhuan Gong, John Jenkins, Eric R. Schendel, Norbert Podhorszki, Qing Liu, Scott Klasky, and Nagiza F. Samatova

Session 4B: Storage and I/O Systems I

HyCache+: Towards Scalable High-Performance Caching Middleware for Parallel File Systems ................................................................................................................................. 267
Dongfang Zhao, Kan Qiao, and Ioan Raicu

A Flexible Framework for Asynchronous in Situ and in Transit Analytics for Scientific Simulations .............................................................................................................................. 277
Matthieu Dreher and Bruno Raffin

Iteration Based Collective I/O Strategy for Parallel I/O Systems .................................................... 287
Zhixiang Wang, Xuanhua Shi, Hai Jin, Song Wu, and Yong Chen

Session 4C: Algorithm

Efficiently Handling Skew in Outer Joins on Distributed Systems .................................................. 295
Long Cheng, Spyros Kotoulas, Tomas E. Ward, and Georgios Theodoropoulos

A PGAS Execution Model for Efficient Stencil Computation on Many-Core Processors .................. 305
Mitsuru Ikei and Mitsuhashi Sato

A Branch-and-Bound Algorithm for Autonomic Adaptation of Multi-cloud Applications .................. 315
André Almeida, Francisco Dantas, Everton Cavalcante, and Thais Batista
**Session 5A: Scheduling**

Decentralized Scheduling and Load Balancing for Parallel Programs .................................................................324

*Gary Jackson, Pete Keleher, and Alan Sussman*

Multi-objective Scheduling for Heterogeneous Server Systems with Machine Placement ........................................334

*Hongyang Sun, Patricia Stolf, Jean-Marc Pierson, and Georges Da Costa*

PLASTiCC: Predictive Look-Ahead Scheduling for Continuous Dataflows on Clouds ........................................344

*Alok Gautam Kumbhare, Yogesh Simmhan, and Viktor K. Prasanna*

Link-Heterogeneous Work Stealing ......................................................................................................................354

*Trong-Tuan Vu and Bilel Derbel*

**Session 5B: Virtual Machines**

Time-Bound, Thread-Based Live Migration of Virtual Machines ........................................................................364

*Kasidit Chanchio and Phitthak Thaenkaew*

Modeling the Virtual Machine Launching Overhead under Fermicloud ............................................................374

*Hao Wu, Shangping Ren, Gabriele Garzoglio, Steven Timm, Gerard Bernabeu, and Seo-Young Noh*

Controlling the Deployment of Virtual Machines on Clusters and Clouds for Scientific Computing in CBRAIN ................384

*Tristan Glatard, Marc-Etienne Rousseau, Pierre Rioux, Reza Adalat, and Alan C. Evans*

MIMP: Deadline and Interference Aware Scheduling of Hadoop Virtual Machines ........................................394

*Wei Zhang, Sundaresan Rajasekaran, Timothy Wood, and Mingfa Zhu*

**Session 5C: Datacenters and Distributed Computing**

Achieving Efficient Distributed Scheduling with Message Queues in the Cloud for Many-Task Computing and High-Performance Computing ................................................................................404

*Iman Sadooghi, Sandeep Palur, Ajay Anthony, Isha Kapur, Karthik Belagodu, Pankaj Purandare, Kiran Ramamurty, Ke Wang, and Ioan Raicu*

Efficient Checkpointing of Virtual Machines Using Virtual Machine Introspection ........................................414

*Ferrol Aderholdt, Fang Han, Stephen L. Scott, and Thomas Naughton*

Analysis of Labor Efforts and their Impact Factors to Solve Server Incidents in Datacenters ..................................424

*Ioana Giurgiu, Jasmina Bogojeska, Sergii Nikolaiev, George Stark, and Dorothea Wiesmann*

CMcloud: Cloud Platform for Cost-Effective Offloading of Mobile Applications ..................................................434

*Dongju Chae, Jihun Kim, Jangwoo Kim, Jong Kim, Seungjun Yang, Yeongpil Cho, Yongin Kwon, and Yunheung Paek*

**Session 6A: Apps II**

A Credential Store for Multi-tenant Science Gateways ........................................................................................445

*Thejaka Amila Kanewala, Suresh Marru, Jim Basney, and Marlon Pierce*

Cluster-Based SNP Calling on Large-Scale Genome Sequencing Data .................................................................455

*Mucahid Kutlu and Gagan Agrawal*
Platform Calibration for Load Balancing of Large Simulations: TLM Case ................................................................. 465

Cristian Ruiz, Mihai Alexandru, Olivier Richard, Thierry Monteil, and Hervé Aubert

Session 6B: Architecture

Enhancing Locality via Caching in the GMU Protocol ........................................................................................................ 473

Hugo Pimentel, Paolo Romano, Sebastiano Peluso, and Pedro Ruivo

Energy-Efficient Collective Reduce and Allreduce Operations on Distributed GPUs ........................................................... 483

Lena Oden, Benjamin Klenk, and Holger Fröning

Network Topology Optimization for Data Aggregation ........................................................................................................ 493

Soham Das and Sartaj Sahni

Session 6C: Storage and I/O Systems II

hatS: A Heterogeneity-Aware Tiered Storage for Hadoop ........................................................................................................ 502

K.R. Krish, Ali Anwar, and Ali R. Butt

A Study of Effective Replica Reconstruction Schemes at Node Deletion for HDFS .............................................................. 512

Asami Higai, Atsuko Takefusa, Hidemoto Nakada, and Masato Oguchi

A Novel Zero-Knowledge Scheme for Proof of Data Possession in Cloud Storage Applications .................................................. 522

Nesrine Kaaniche, Ethmane El Moustaine, and Maryline Laurent

Poster Papers

A High Efficient Disk Scheduling Framework with QoS Mechanism in Xen-Based Cloud Platforms .................................................. 532

Tseng-Yi Chen, Hsin-Wen Wei, Ying-Jie Chen, Nia-Yuan Chang, Tsan-Sheng Hsu, and Wei-Kuan Shih

Demo Paper: Automatic Provisioning, Deploy and Monitoring of Virtual Machines Based on Security Service Level Agreement in the Cloud ........................................................................................................ 536

Kazi Wali Ullah and Abu Shoheil Ahmed

RESeED: A Tool for Regular Expression Search over Encrypted Data in Cloud Storage ...................................................... 538

Mohsen Amini Salehi, Thomas Caldwell, Alejandro Fernandez, Emmanuel Mickiewicz, Eric W.D. Rozier, Saman Zonouz, and David Redberg

Wave: Trigger Based Synchronous Data Process System ........................................................................................................ 540

Kun Lu, Mingming Sun, Changlong Li, Hang Zhuang, Jinhong Zhou, and Xuehai Zhou

Network Traffic-Aware Virtual Machine Placement with Availability Guarantees Based on Shadows .................................................. 542

Qian Zhang, Mingyu Li, and Xiaohui Hu

An Architecture for Orchestrating Hadoop Applications in Hybrid Cloud .................................................................................. 544

Carlos R. Senna, Luis G.C. Russi, and Edmundo R.M. Madeira

Performance Evaluation of an IaaS Opportunistic Cloud Computing ...................................................................................... 546

Cesar O. Diaz, Johnatan E. Pecero, Pascal Bouvry, German Sotelo, Mario Villamizar, and Harold Castro
Expanding Tasks of Logical Workflows Into Independent Workflows for Improved Scalability .................................................................548
Nicholas Hazekamp, Olivia Choudhury, Sandra Gesing, Scott Emrich, and Douglas Thain

Provenance-Based Prediction Scheme for Object Storage System in HPC ........................................................................550
Dong Dai, Yong Chen, Dries Kimpe, and Rob Ross

An Adaptive Separation-Aware FTL for Improving the Efficiency of Garbage Collection in SSDs ..........................................................552
Wei Xie and Yong Chen

Doctoral Symposium

Strategy-Proof Mechanisms for Resource Management in Clouds ..................................................................................554
Lena Mashayekhy and Daniel Grosu

Proactive Workload Consolidation for Reducing Energy Cost over a Given Time Horizon ..................................................................................558
Milan De Cauwer, Deepak Mehta, Barry O’Sullivan, Helmut Simonis, and Hadrien Cambazard

Runtime Adaptation for Autonomic Heterogeneous Computing ..................................................................................562
Thomas R.W. Scogland and Wu-Chun Feng

Towards Generic Metadata Management in Distributed Science Gateway Infrastructures ......................................................566
Richard Grunzke, René Jäkel, Wolfgang E. Nagel, and Sandra Gesing

Compiler Optimization for Extreme-Scale Scripting ..................................................................................571
Timothy G. Armstrong, Justin M. Wozniak, Michael Wilde, and Ian T. Foster

Metrics, Models and Methodologies for Energy-Proportional Computing ..................................................................................575
Balaji Subramaniam

Parallel Computing with P2P Desktop Grids ..................................................................................579
Gary Jackson

Advanced Virtualization Techniques for High Performance Cloud Cyberinfrastructure ......................................................583
Andrew J. Younge and Geoffrey C. Fox

Supporting Queries and Analyses of Large-Scale Social Media Data with Customizable and Scalable Indexing Techniques over NoSQL Databases ..................................................................................587
Xiaoming Gao and Judy Qiu

Towards an SLA-Based Service Allocation in Multi-cloud Environments ..................................................................................591
Soodeh Farokhi

Workshop Papers
Extreme Green and Energy Efficiency in Large Scale Distributed Systems (ExtremeGreen)

Energy-Aware Profit Maximizing Scheduling Algorithm for Heterogeneous Computing Systems ................................................................. Kyle M. Tarplee, Anthony A. Maciejewski, and Howard Jay Siegel

Energy Consumption of Photo Sharing in Online Social Networks ........................................................................................................ Fatemeh Jalali, Chrispin Gray, Arun Vishwanath, Robert Ayre, Tansu Alpcan, Kerry Hinton, and Rodney S. Tucker

Power Consumption Evaluation of an MHD Simulation with CPU Power Capping .................................................................................. Keiichiro Fukazawa, Masatsugu Ueda, Mutsumi Aoyagi, Tomonori Tsuchita, Kyohei Yoshida, Aruta Uehara, Masakazu Kuze, Yuichi Inadomi, and Koji Inoue

A Game-Theoretic Approach to Coalition Formation in Green Cloud Federations ...................................................................................... Marco Guazzone, Cosimo Anglano, and Matteo Sereno

Energy-Aware Data Transfer Tuning ..................................................................................................................................................... Ismail Alan, Engin Arslan, and Tevfik Kosar

The Third Workshop on Data-Intensive Process Management in Large-Scale Sensor Systems (DPMSS)

Fuzzy Assisted Event Driven Data Collection from Sensor Nodes in Sensor-Cloud Infrastructure .................................................................................. Suman Sankar Bhunia, Jayita Pal, and Nandini Mukherjee

Cloud Supported Building Data Analytics ........................................................................................................................................ Ioan Petri, Omer Rana, Yacine Rezgui, Haijiang Li, Tom Beach, Mengsong Zou, Javier Diaz-Montes, and Manish Parashar

Enforcing Quality of Service on OpenNebula-Based Shared Clouds ........................................................................................................... Rafael Tolosana-Calasanz, José Ángel Bañares, Omer Rana, Congduc Pham, Erotokritos Xydas, Charalampos Marmaras, Panagiotis Papadopoulos, and Liana Cipcigan


A Cloud-Based Framework for Supporting Effective and Efficient OLAP in Big Data Environments ......................................................................................... Alfredo Cuzzocrea and Rim Moussa

Multiobjective Communication Optimization for Cloud-Integrated Body Sensor Networks ........................................................................ Dung H. Phan, Junichi Suzuki, Shingo Omura, Katsuya Oba, and Athanasios Vasilakos

First International Workshop on Cloud for Bio (C4BIO 2014)

From Scripted HPC-Based NGS Pipelines to Workflows on the Cloud ................................................................................................... Jacek Cała, Yaobo Xu, Eldarina Azfar Wijaya, and Paolo Missier
A Survey of Approaches and Frameworks to Carry Out Genomic Data Analysis on the Cloud ...............................................................701
Philip C. Church and Andrzej Goscinski

Accelerating Comparative Genomics Workflows in a Distributed Environment with Optimized Data Partitioning ...............................................................711
Olivia Choudhury, Nicholas L. Hazekamp, Douglas Thain, and Scott Emrich

Integration of Clustering and Multidimensional Scaling to Determine Phylogenetic Trees as Spherical Phylogenograms Visualized in 3 Dimensions .................................................................................720
Yang Ruan, Geoffrey L. House, Saliya Ekanayake, Ursel Schütte, James D. Bever, Haixu Tang, and Geoffrey Fox

A Performance Evaluation of Sequence Alignment Software in Virtualized Environments ...........................................................................................................730
Zachary J. Estrada, Zachary Stephens, Cuong Pham, Zbigniew Kalbarczyk, and Ravishankar K. Iyer

A Storage Policy for a Hybrid Federated Cloud platform: A Case Study for Bioinformatics ........................................................................................................738
Deric Lima, Breno Moura, Gabriel Oliveira, Edward Ribeiro, Aleteia Araujo, Maristela Holanda, Roberto Togawa, and Maria Emilia Walter

Evaluation of the Feasibility of Making Large-Scale X-Ray Tomography Reconstructions on Clouds ..........................................................................................748
Estefania Serrano, Guzman Bermejo, Javier Garcia Blas, and Jesus Carretero

2014 Workshop on Clusters, Clouds and Grids for Health (CCGrid-Health)

Global Initiative for Sentinel e-Health Network on Grid (GINSENG): Medical Data Integration and Semantic Developments for Epidemiology ............................................................................755
Sébastien Cipière, Guillaume Ereteo, Alban Gaignard, Nouha Boujelben, Sébastien Gaspar, Vincent Breton, Fréséric Cervenansky, David R.C. Hill, Tristan Glatard, David Manset, Johan Montagnat, Jérôme Revillard, and Lydia Maigne

Extending XNAT towards a Cloud-Based Quality Assessment Platform for Retinal Optical Coherence Tomographies ........................................................................764
Jie Wu, Christoph Jansen, Maximilian Beier, Michael Witt, and Dagmar Krefting

Distributed Detection of Cancer Cells in High-Throughput Cellular Spike Streams ..................................................................................................................774
Abdul Hafeez, M. Mustafa Rafique, and Ali R. Butt

Second International Workshop on Assured Cloud Computing (WACC 2014)

A High Performance, QoS-Enabled, S3-Based Object Store .................................................................................................................................784
Yusuke Tanimura, Seiya Yanagita, and Takahiro Hamanishi

Towards Cloud, Service and Tenant Classification for Cloud Computing ..................................................................................................................792
Sebastian Jeuk, Jakub Szefer, and Shi Zhou
On the Evaluation of VM Provisioning Time in Cloud Platforms for Mission-Critical Infrastructures
Gabriella Carrozza, Luigi Battaglia, Vittorio Manetti, Antonio Marotta, Roberto Canonico, and Stefano Avallone

2014 International Workshop on Data Vitalization and Universal Village: Extended Version of Smart Cities (DV&UV)

Pragmatic Oriented Data Interoperability for Smart Healthcare Information Systems
Shixiong Liu, Weizi Li, and Kecheng Liu

Geospatial Sensor Web Resource Management System for Smart City: Design and Implementation
Jia Li and Nengcheng Chen

A Robust and Fast Reconstruction Framework for Noisy and Large Point Cloud Data
Xiang Feng, Xiaoqing Yu, Wanggen Wan, Fabien Pfaender, and J. Alfredo Sánchez

A Survey on Workflow Management and Scheduling in Cloud Computing
Li Liu, Miao Zhang, Yuqing Lin, and Liangjuan Qin

Intelligent Congestion Avoidance Algorithm and System—Application of Data Vitalization
Yan Huang, Hao Sheng, and Jiahui Chen

Variable Window for Outlier Detection and Impulsive Noise Recognition in Range Images
Jian Wang, Lin Mei, Yi Li, Jian-Ye Li, Kun Zhao, and Yuan Yao

Data Vitalization's Perspective Towards Smart City: A Reference Model for Data Service Oriented Architecture
Zhang Xiong, Yanwei Zheng, and Chao Li

First International Workshop on Scalable Computing for Real-Time Big Data Applications (SCRAMBL)

Scalable Infrastructures for Data in Motion
David Ediger, Rob McColl, Jason Poovey, and Dan Campbell

Towards In-Order and Exactly-Once Delivery Using Hierarchical Distributed Message Queues
Dharmit Patel, Faraj Khasib, Iman Sadooghi, and Ioan Raicu

A Scalable System for Community Discovery in Twitter During Hurricane Sandy
Yin Huang, Han Dong, Yelena Yesha, and Shujia Zhou

A Scalable Real-Time Photometric System for Automatic Astronomical Observations on Dome A
Ce Yu, Lianmeng Li, Jizhou Sun, Jian Xiao, Jiajun Li, and Zhaohui Shang
Cloud for Business, Industry and Enterprises (C4BIE 2014)
Improving Resource Matchmaking through Feedback Integration .................................................................906
  Christian Haas, Ioan Petri, and Omer Rana
MO-BIZZ: Fostering Mobile Business through Enhanced Cloud Solutions .......................................................915
  Alexander Stanik, Odej Kao, Rui Martins, António Cruz, and Dimitrios Tektonidis
A Comparative Study of Predictive Models for Cloud Infrastructure Management ...........................................923
  Mahesh Balaji, G. Subrahmanya VRK Rao, and Ch. Aswani Kumar

SCALE Challenge
V for Vicissitude: The Challenge of Scaling Complex Big Data Workflows .........................................................927
  Bogdan Ghiț, Mihai Capotă, Tim Hegeman, Jan Hidders, Dick Epema, and Alexandru Iosup
Emulation at Very Large Scale with Distem ........................................................................................................933
  Tomasz Buchert, Emmanuel Jeanvoine, and Lucas Nussbaum
ToMaR—A Data Generator for Large Volumes of Content .................................................................................937
  Rainer Schmidt, Matthias Rella, and Sven Schlarb
Exploring Infiniband Hardware Virtualization in OpenNebula towards Efficient High-Performance Computing .................................................................943
  Tiago Pais Pitta de Lacerda Ruivo, Gerard Bernabeu Altayo, Gabriele Garzoglio,
  Steven Timm, Hyun Woo Kim, Seo-Young Noh, and Ioan Raicu

Author Index .................................................................................................................................................................949