

2014 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing

CCGrid 2014

Table of Contents

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

Session 1: Best Papers

Improving I/O Throughput of Scientific Applications Using Transparent Parallel Compression	1
<i>Tekin Bicer, Jian Yin, and Gagan Agrawal</i>	
Performance Models for CPU-GPU Data Transfers	11
<i>B. van Werkhoven, J. Maassen, F.J. Seinstra, and H.E. Bal</i>	
A User-Level InfiniBand-Based File System and Checkpoint Strategy for Burst Buffers	21
<i>Kento Sato, Kathryn Mohror, Adam Moody, Todd Gamblin, Bronis R. de Supinski, Naoya Maruyama, and Satoshi Matsuoka</i>	

Session 2A: MapReduce

Tagged-MapReduce: A General Framework for Secure Computing with Mixed-Sensitivity Data on Hybrid Clouds	31
<i>Chunwang Zhang, Ee-Chien Chang, and Roland H.C. Yap</i>	
Toward Detecting Compromised MapReduce Workers through Log Analysis	41
<i>Eunjung Yoon and Anna Squicciarini</i>	
MapReduce Analysis for Cloud-Archived Data	51
<i>Balaji Palanisamy, Aameek Singh, Nagapramod Mandagere, Gabriel Alatorre, and Ling Liu</i>	
Adaptive MapReduce Scheduling in Shared Environments	61
<i>Jordà Polo, Yolanda Becerra, David Carrera, Jordi Torres, Eduard Ayguadé, and Malgorzata Steinder</i>	

Session 2B: Energy and the Environment

Enabling Efficient Power Provisioning for Enterprise Applications	71
<i>Balaji Subramaniam and Wu-Chun Feng</i>	
Analytical/ML Mixed Approach for Concurrency Regulation in Software Transactional Memory	81
<i>Diego Ruggetti, Pierangelo Di Sanzo, Bruno Ciciani, and Francesco Quaglia</i>	
Bridging Data in the Clouds: An Environment-Aware System for Geographically Distributed Data Transfers	92
<i>Radu Tudoran, Alexandru Costan, Rui Wang, Luc Bougé, and Gabriel Antoniu</i>	
Cost-Efficient, Reliable, Utility-Based Session Management in the Cloud	102
<i>Benjamin Byholm and Iván Porres</i>	

Session 2C: Resource Management

A Workflow-Inspired, Modular and Robust Approach to Experiments in Distributed Systems	112
<i>Tomasz Buchert, Lucas Nussbaum, and Jens Gustedt</i>	
Resource Usage Control in Multi-tenant Applications	122
<i>Rouven Krebs, Simon Spinner, Nadia Ahmed, and Samuel Kounev</i>	
Mapping Algorithms Optimizing the Overall Manhattan Distance for Pre-Occupied Cluster Computers in SLA-Based Grid Environments	132
<i>Barry Linnert, Joerg Schneider, and Lars-Olof Burchard</i>	
SLA-Based Profit Optimization in Cloud Bursting PaaS	141
<i>Djawida Dib, Nikos Parlavantzas, and Christine Morin</i>	

Session 3A: Apps I

Evaluating Streaming Strategies for Event Processing Across Infrastructure Clouds	151
<i>Radu Tudoran, Kate Keahey, Pierre Riteau, Sergey Panitkin, and Gabriel Antoniu</i>	
CUDAAlign 3.0: Parallel Biological Sequence Comparison in Large GPU Clusters	160
<i>Edans F. de O. Sandes, Guillermo Miranda, Alba C.M.A. de Melo, Xavier Martorell, and Eduard Ayguadé</i>	
Opportunistic High Energy Physics Computing in User Space with Parrot	170
<i>Dillon Skeeahan, Paul Brenner, Ben Tovar, Douglas Thain, N. Valls, A. Woodard, M. Wolf, T. Pearson, S. Lynch, and K. Lannon</i>	

Session 3B: Message Passing

Towards an MPI-Like Framework for the Azure Cloud Platform	176
<i>Dinesh Agarwal, Sara Karamati, Satish Puri, and Sushil K. Prasad</i>	
NoCMsg: Scalable NoC-Based Message Passing	186
<i>Christopher Zimmer and Frank Mueller</i>	
Modeling and Optimizing Large-Scale Wide-Area Data Transfers	196
<i>Rajkumar Kettimuthu, Gayane Vardoyan, Gagan Agrawal, and P. Sadayappan</i>	

Session 3C: Elasticity and Adaptation

A Language Support for Cloud Elasticity Management	206
<i>Yousri Kouki, Frederico Alvares de Oliveira Jr., Simon Dupont, and Thomas Ledoux</i>	
Elastic MapReduce Execution	216
<i>Wei Xiang Goh and Kian-Lee Tan</i>	
JCatascopia: Monitoring Elastically Adaptive Applications in the Cloud	226
<i>Demetris Trihinas, George Pallis, and Marios D. Dikaiakos</i>	

Session 4A: Big Data

Towards a Collective Layer in the Big Data Stack	236
<i>Thilina Gunarathne, Judy Qiu, and Dennis Gannon</i>	
Flexpath: Type-Based Publish/Subscribe System for Large-Scale Science Analytics	246
<i>Jai Dayal, Drew Bratcher, Greg Eisenhauer, Karsten Schwan, Matthew Wolf, Xuechen Zhang, Hasan Abbasi, Scott Klasky, and Norbert Podhorszki</i>	
Transparent in Situ Data Transformations in ADIOS	256
<i>David A. Boyuka, Sriram Lakshminarasimham, Xiaocheng Zou, Zhenhuan Gong, John Jenkins, Eric R. Schendel, Norbert Podhorszki, Qing Liu, Scott Klasky, and Nagiza F. Samatova</i>	

Session 4B: Storage and I/O Systems I

HyCache+: Towards Scalable High-Performance Caching Middleware for Parallel File Systems	267
<i>Dongfang Zhao, Kan Qiao, and Ioan Raicu</i>	
A Flexible Framework for Asynchronous in Situ and in Transit Analytics for Scientific Simulations	277
<i>Matthieu Dreher and Bruno Raffin</i>	
Iteration Based Collective I/O Strategy for Parallel I/O Systems	287
<i>Zhixiang Wang, Xuanhua Shi, Hai Jin, Song Wu, and Yong Chen</i>	

Session 4C: Algorithm

Efficiently Handling Skew in Outer Joins on Distributed Systems	295
<i>Long Cheng, Spyros Kotoulas, Tomas E. Ward, and Georgios Theodoropoulos</i>	
A PGAS Execution Model for Efficient Stencil Computation on Many-Core Processors	305
<i>Mitsuru Ikei and Mitsuhsa Sato</i>	
A Branch-and-Bound Algorithm for Autonomic Adaptation of Multi-cloud Applications	315
<i>André Almeida, Francisco Dantas, Everton Cavalcante, and Thais Batista</i>	

Session 5A: Scheduling

Decentralized Scheduling and Load Balancing for Parallel Programs	324
<i>Gary Jackson, Pete Keleher, and Alan Sussman</i>	
Multi-objective Scheduling for Heterogeneous Server Systems with Machine Placement	334
<i>Hongyang Sun, Patricia Stolf, Jean-Marc Pierson, and Georges Da Costa</i>	
PLASiCC: Predictive Look-Ahead Scheduling for Continuous Dataflows on Clouds	344
<i>Alok Gautam Kumbhare, Yogesh Simmhan, and Viktor K. Prasanna</i>	
Link-Heterogeneous Work Stealing	354
<i>Trong-Tuan Vu and Bilel Derbel</i>	

Session 5B: Virtual Machines

Time-Bound, Thread-Based Live Migration of Virtual Machines	364
<i>Kasidit Chanchio and Phithak Thaenkaew</i>	
Modeling the Virtual Machine Launching Overhead under Fermicloud	374
<i>Hao Wu, Shangping Ren, Gabriele Garzoglio, Steven Timm, Gerard Bernabeu, and Seo-Young Noh</i>	
Controlling the Deployment of Virtual Machines on Clusters and Clouds for Scientific Computing in CBRAIN	384
<i>Tristan Glatard, Marc-Etienne Rousseau, Pierre Rioux, Reza Adalat, and Alan C. Evans</i>	
MIMP: Deadline and Interference Aware Scheduling of Hadoop Virtual Machines	394
<i>Wei Zhang, Sundaresan Rajasekaran, Timothy Wood, and Mingfa Zhu</i>	

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
<i>Iman Sadooghi, Sandeep Palur, Ajay Anthony, Isha Kapur, Karthik Belagodu, Pankaj Purandare, Kiran Ramamurty, Ke Wang, and Ioan Raicu</i>	
Efficient Checkpointing of Virtual Machines Using Virtual Machine Introspection	414
<i>Ferrol Aderholdt, Fang Han, Stephen L. Scott, and Thomas Naughton</i>	
Analysis of Labor Efforts and their Impact Factors to Solve Server Incidents in Datacenters	424
<i>Ioana Giurgiu, Jasmina Bogojeska, Sergii Nikolaiev, George Stark, and Dorothea Wiesmann</i>	
CMcloud: Cloud Platform for Cost-Effective Offloading of Mobile Applications	434
<i>Dongju Chae, Jihun Kim, Jangwoo Kim, Jong Kim, Seungjun Yang, Yeongpil Cho, Yongin Kwon, and Yunheung Paek</i>	

Session 6A: Apps II

A Credential Store for Multi-tenant Science Gateways	445
<i>Thejaka Amila Kanewala, Suresh Marru, Jim Basney, and Marlon Pierce</i>	
Cluster-Based SNP Calling on Large-Scale Genome Sequencing Data	455
<i>Mucahid Kutlu and Gagan Agrawal</i>	

Platform Calibration for Load Balancing of Large Simulations: TLM Case	465
<i>Cristian Ruiz, Mihai Alexandru, Olivier Richard, Thierry Monteil, and Hervé Aubert</i>	

Session 6B: Architecture

Enhancing Locality via Caching in the GMU Protocol	473
<i>Hugo Pimentel, Paolo Romano, Sebastiano Peluso, and Pedro Ruivo</i>	
Energy-Efficient Collective Reduce and Allreduce Operations on Distributed GPUs	483
<i>Lena Oden, Benjamin Klenk, and Holger Fröning</i>	
Network Topology Optimization for Data Aggregation	493
<i>Soham Das and Sartaj Sahni</i>	

Session 6C: Storage and I/O Systems II

hatS: A Heterogeneity-Aware Tiered Storage for Hadoop	502
<i>K.R. Krish, Ali Anwar, and Ali R. Butt</i>	
A Study of Effective Replica Reconstruction Schemes at Node Deletion for HDFS	512
<i>Asami Higai, Atsuko Takefusa, Hidemoto Nakada, and Masato Oguchi</i>	
A Novel Zero-Knowledge Scheme for Proof of Data Possession in Cloud Storage Applications	522
<i>Nesrine Kaaniche, Ethmane El Moustaine, and Maryline Laurent</i>	

Poster Papers

A High Efficient Disk Scheduling Framework with QoS Mechanism in Xen-Based Cloud Platforms	532
<i>Tseng-Yi Chen, Hsin-Wen Wei, Ying-Jie Chen, Nia-Yuan Chang, Tsan-Sheng Hsu, and Wei-Kuan Shih</i>	
Demo Paper: Automatic Provisioning, Deploy and Monitoring of Virtual Machines Based on Security Service Level Agreement in the Cloud	536
<i>Kazi Wali Ullah and Abu Shohel Ahmed</i>	
RESeED: A Tool for Regular Expression Search over Encrypted Data in Cloud Storage	538
<i>Mohsen Amini Salehi, Thomas Caldwell, Alejandro Fernandez, Emmanuel Mickiewicz, Eric W.D. Rozier, Saman Zonouz, and David Redberg</i>	
Wave: Trigger Based Synchronous Data Process System	540
<i>Kun Lu, Mingming Sun, Changlong Li, Hang Zhuang, Jinhong Zhou, and Xuehai Zhou</i>	
Network Traffic-Aware Virtual Machine Placement with Availability Guarantees Based on Shadows	542
<i>Qian Zhang, Mingyu Li, and Xiaohui Hu</i>	
An Architecture for Orchestrating Hadoop Applications in Hybrid Cloud	544
<i>Carlos R. Senna, Luis G.C. Russi, and Edmundo R.M. Madeira</i>	
Performance Evaluation of an IaaS Opportunistic Cloud Computing	546
<i>Cesar O. Diaz, Johnatan E. Pecero, Pascal Bouvry, German Sotelo, Mario Villamizar, and Harold Castro</i>	

Expanding Tasks of Logical Workflows Into Independent Workflows for Improved Scalability	548
<i>Nicholas Hazekamp, Olivia Choudhury, Sandra Gesing, Scott Emrich, and Douglas Thain</i>	
Provenance-Based Prediction Scheme for Object Storage System in HPC	550
<i>Dong Dai, Yong Chen, Dries Kimpe, and Rob Ross</i>	
An Adaptive Separation-Aware FTL for Improving the Efficiency of Garbage Collection in SSDs	552
<i>Wei Xie and Yong Chen</i>	

Doctoral Symposium

Strategy-Proof Mechanisms for Resource Management in Clouds	554
<i>Lena Mashayekhy and Daniel Grosu</i>	
Proactive Workload Consolidation for Reducing Energy Cost over a Given Time Horizon	558
<i>Milan De Cauwer, Deepak Mehta, Barry O'Sullivan, Helmut Simonis, and Hadrien Cambazard</i>	
Runtime Adaptation for Autonomic Heterogeneous Computing	562
<i>Thomas R.W. Scogland and Wu-Chun Feng</i>	
Towards Generic Metadata Management in Distributed Science Gateway Infrastructures	566
<i>Richard Grunzke, René Jäkel, Wolfgang E. Nagel, and Sandra Gesing</i>	
Compiler Optimization for Extreme-Scale Scripting	571
<i>Timothy G. Armstrong, Justin M. Wozniak, Michael Wilde, and Ian T. Foster</i>	
Metrics, Models and Methodologies for Energy-Proportional Computing	575
<i>Balaji Subramaniam</i>	
Parallel Computing with P2P Desktop Grids	579
<i>Gary Jackson</i>	
Advanced Virtualization Techniques for High Performance Cloud Cyberinfrastructure	583
<i>Andrew J. Younge and Geoffrey C. Fox</i>	
Supporting Queries and Analyses of Large-Scale Social Media Data with Customizable and Scalable Indexing Techniques over NoSQL Databases	587
<i>Xiaoming Gao and Judy Qiu</i>	
Towards an SLA-Based Service Allocation in Multi-cloud Environments	591
<i>Soodeh Farokhi</i>	

Workshop Papers

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

Energy-Aware Profit Maximizing Scheduling Algorithm for Heterogeneous Computing Systems	595
<i>Kyle M. Tarplee, Anthony A. Maciejewski, and Howard Jay Siegel</i>	
Energy Consumption of Photo Sharing in Online Social Networks	604
<i>Fatemeh Jalali, Chrispin Gray, Arun Vishwanath, Robert Ayre, Tansu Alpcan, Kerry Hinton, and Rodney S. Tucker</i>	
Power Consumption Evaluation of an MHD Simulation with CPU Power Capping	612
<i>Keiichiro Fukazawa, Masatsugu Ueda, Mutsumi Aoyagi, Tomonori Tsuhata, Kyohei Yoshida, Aruta Uehara, Masakazu Kuze, Yuichi Inadomi, and Koji Inoue</i>	
A Game-Theoretic Approach to Coalition Formation in Green Cloud Federations	618
<i>Marco Guazzzone, Cosimo Anglano, and Matteo Sereno</i>	
Energy-Aware Data Transfer Tuning	626
<i>Ismail Alan, Engin Arslan, and Tevfik Kosar</i>	

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	635
<i>Suman Sankar Bhunia, Jayita Pal, and Nandini Mukherjee</i>	
Cloud Supported Building Data Analytics	641
<i>Ioan Petri, Omer Rana, Yacine Rezgui, Haijiang Li, Tom Beach, Mengsong Zou, Javier Diaz-Montes, and Manish Parashar</i>	
Enforcing Quality of Service on OpenNebula-Based Shared Clouds	651
<i>Rafael Tolosana-Calasanz, José Ángel Bañares, Omer Rana, Congduc Pham, Erotokritos Xydias, Charalampos Marmaras, Panagiotis Papadopoulos, and Liana Cipcigan</i>	
A New Parallelism-Capable Clustering Algorithm for Wireless Sensor Networks	660
<i>Alireza T. Boloorch, M.H. Samadzadeh, and Nazarin Rahnavard</i>	
An Evaluation Framework for Buildings-Oriented Wireless Sensor Networks	670
<i>Antonio Guerrieri, Giancarlo Fortino, and Wilma Russo</i>	
A Cloud-Based Framework for Supporting Effective and Efficient OLAP in Big Data Environments	680
<i>Alfredo Cuzzocrea and Rim Moussa</i>	
Multiobjective Communication Optimization for Cloud-Integrated Body Sensor Networks	685
<i>Dung H. Phan, Junichi Suzuki, Shingo Omura, Katsuya Oba, and Athanasios Vasilakos</i>	

First International Workshop on Cloud for Bio (C4BIO 2014)

From Scripted HPC-Based NGS Pipelines to Workflows on the Cloud	694
<i>Jacek Cala, Yaobo Xu, Eldarina Azfar Wijaya, and Paolo Missier</i>	

A Survey of Approaches and Frameworks to Carry Out Genomic Data Analysis on the Cloud	701
<i>Philip C. Church and Andrzej Goscinski</i>	
Accelerating Comparative Genomics Workflows in a Distributed Environment with Optimized Data Partitioning	711
<i>Olivia Choudhury, Nicholas L. Hazekamp, Douglas Thain, and Scott Emrich</i>	
Integration of Clustering and Multidimensional Scaling to Determine Phylogenetic Trees as Spherical Phylograms Visualized in 3 Dimensions	720
<i>Yang Ruan, Geoffrey L. House, Saliya Ekanayake, Ursel Schütte, James D. Bever, Haixu Tang, and Geoffrey Fox</i>	
A Performance Evaluation of Sequence Alignment Software in Virtualized Environments	730
<i>Zachary J. Estrada, Zachary Stephens, Cuong Pham, Zbigniew Kalbarczyk, and Ravishankar K. Iyer</i>	
A Storage Policy for a Hybrid Federated Cloud platform: A Case Study for Bioinformatics	738
<i>Deric Lima, Breno Moura, Gabriel Oliveira, Edward Ribeiro, Aleteia Araujo, Maristela Holanda, Roberto Togawa, and Maria Emilia Walter</i>	
Evaluation of the Feasibility of Making Large-Scale X-Ray Tomography Reconstructions on Clouds	748
<i>Estefania Serrano, Guzman Bermejo, Javier Garcia Blas, and Jesus Carretero</i>	

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
<i>Sébastien Cipièrre, Guillaume Ereteo, Alban Gaignard, Nouha Boujelben, Sébastien Gaspard, Vincent Breton, Fréséric Cervenansky, David R.C. Hill, Tristan Glatard, David Manset, Johan Montagnat, Jérôme Revillard, and Lydia Maigne</i>	
Extending XNAT towards a Cloud-Based Quality Assessment Platform for Retinal Optical Coherence Tomographies	764
<i>Jie Wu, Christoph Jansen, Maximilian Beier, Michael Witt, and Dagmar Krefting</i>	
Distributed Detection of Cancer Cells in High-Throughput Cellular Spike Streams	774
<i>Abdul Hafeez, M. Mustafa Rafique, and Ali R. Butt</i>	

Second International Workshop on Assured Cloud Computing (WACC 2014)

A High Performance, QoS-Enabled, S3-Based Object Store	784
<i>Yusuke Tanimura, Seiya Yanagita, and Takahiro Hamanishi</i>	
Towards Cloud, Service and Tenant Classification for Cloud Computing	792
<i>Sebastian Jeuk, Jakub Szefer, and Shi Zhou</i>	

On the Evaluation of VM Provisioning Time in Cloud Platforms for Mission-Critical Infrastructures	802
<i>Gabriella Carrozza, Luigi Battaglia, Vittorio Manetti, Antonio Marotta, Roberto Canonico, and Stefano Avallone</i>	

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	811
<i>Shixiong Liu, Weizi Li, and Kecheng Liu</i>	
Geospatial Sensor Web Resource Management System for Smart City: Design and Implementation	819
<i>Jia Li and Nengcheng Chen</i>	
A Robust and Fast Reconstruction Framework for Noisy and Large Point Cloud Data	828
<i>Xiang Feng, Xiaoqing Yu, Wanggen Wan, Fabien Pfaënder, and J. Alfredo Sánchez</i>	
A Survey on Workflow Management and Scheduling in Cloud Computing	837
<i>Li Liu, Miao Zhang, Yuqing Lin, and Liangjuan Qin</i>	
Intelligent Congestion Avoidance Algorithm and System—Application of Data Vitalization	847
<i>Yan Huang, Hao Sheng, and Jiahui Chen</i>	
Variable Window for Outlier Detection and Impulsive Noise Recognition in Range Images	857
<i>Jian Wang, Lin Mei, Yi Li, Jian-Ye Li, Kun Zhao, and Yuan Yao</i>	
Data Vitalization's Perspective Towards Smart City: A Reference Model for Data Service Oriented Architecture	865
<i>Zhang Xiong, Yanwei Zheng, and Chao Li</i>	

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

Scalable Infrastructures for Data in Motion	875
<i>David Ediger, Rob McColl, Jason Poovey, and Dan Campbell</i>	
Towards In-Order and Exactly-Once Delivery Using Hierarchical Distributed Message Queues	883
<i>Dharmit Patel, Faraj Khasib, Iman Sadooghi, and Ioan Raicu</i>	
A Scalable System for Community Discovery in Twitter During Hurricane Sandy	893
<i>Yin Huang, Han Dong, Yelena Yesha, and Shujia Zhou</i>	
A Scalable Real-Time Photometric System for Automatic Astronomical Observations on Dome A	900
<i>Ce Yu, Lianmeng Li, Jizhou Sun, Jian Xiao, Jiajun Li, and Zhaohui Shang</i>	

Cloud for Business, Industry and Enterprises (C4BIE 2014)

Improving Resource Matchmaking through Feedback Integration	906
<i>Christian Haas, Ioan Petri, and Omer Rana</i>	
MO-BIZZ: Fostering Mobile Business through Enhanced Cloud Solutions	915
<i>Alexander Stanik, Odej Kao, Rui Martins, António Cruz, and Dimitrios Tektonidis</i>	
A Comparative Study of Predictive Models for Cloud Infrastructure Management	923
<i>Mahesh Balaji, G. Subrahmanya VRK Rao, and Ch. Aswani Kumar</i>	

SCALE Challenge

V for Vicissitude: The Challenge of Scaling Complex Big Data Workflows	927
<i>Bogdan Ghiț, Mihai Capotă, Tim Hegeman, Jan Hidders, Dick Epema, and Alexandru Iosup</i>	
Emulation at Very Large Scale with Distem	933
<i>Tomasz Buchert, Emmanuel Jeanvoine, and Lucas Nussbaum</i>	
ToMaR—A Data Generator for Large Volumes of Content	937
<i>Rainer Schmidt, Matthias Rella, and Sven Schlarb</i>	
Exploring Infiniband Hardware Virtualization in OpenNebula towards Efficient High-Performance Computing	943
<i>Tiago Pais Pitta De Lacerda Ruivo, Gerard Bernabeu Altayo, Gabriele Garzoglio, Steven Timm, Hyun Woo Kim, Seo-Young Noh, and Ioan Raicu</i>	
Author Index	949