# ACM DataCloud-SC11 2011 Concluding Remarks

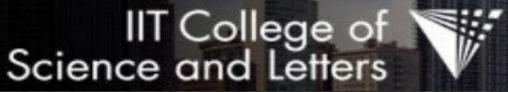
# Chairs

Ioan Raicu, Illinois Institute of Technology & Argonne National Laboratory
Tevfik Kosar, University of Buffalo (SUNY)
Roger Barga, Microsoft Research

ACM DataCloud-SC11 2011 November 14th, 2011

# Sponsorship

# **COMPUTER SCIENCE**





Data-Intensive Distributed
Systems Laboratory



ILLINOIS INSTITUTE OF TECHNOLOGY





The State University of **New York** 

# MTAGS 2011

4th ACM Workshop on Many-Task Computing on Grids and Supercomputers http://datasys.cs.iit.edu/events/MTAGS11/

# **Keynote & Panelists**



David Abramson
Computer Science
Monash e-Research Centre
Monash University
Australia



Dan Reed Corporate Vice President Technology Strategy and Policy and Extreme Computing Microsoft Research



Jack Dongarra
Computer Science Department
University of Tennessee
Computer Science and Math. Div.
Oak Ridge National Laboratory



Daniel S. Katz
Computation Institute
University of Chicago
Math and Computer Science Div
Argonne National Laboratory

			THE CONTRACTOR OF THE CONTRACT
Time	Topic	Authors/Speakers	Affiliation
10:00AM	Introduction from the Chairs	Ioan Raicu	Illinois Institute of Technology & Argonne National Lab.
10.00/ ((1)	\$4.40 (MATERIAL POLICIA)	Yong Zhao	University of Electronic Science and Technology of China
10:10AM	Keynote: Mixing Cloud and Grid Resources for Many Task Computing	David Abramson	Monash University
11:00AM	Panel:	Dan Reed David Abramson	Microsoft Research Monash University
		Jack Dongarra	University of Tennessee & Oak Ridge National Laboratory
	Many-Task Computing meets Exascales	Daniel S. Katz	University of Chicago & Argonne National Laboratory
12:00PM	Lunch 27		
1:30PM	Parallel High-resolution Climate Data Analysis using Swift Best	Matthew Woitaszek	UCAR
		John Dennis	UCAR
	Paper Award Finalist (n)	Taleena Sines	Frostburg State University
2:00PM		Wael Elwasif	Oak Ridge National Laboratory
	A Dependency-Driven Formulation of Parareal: Parallel-in-Time	Samantha Foley	Oak Ridge National Laboratory
		David Bernholdt	Oak Ridge National Laboratory
	Solution of PDEs as a Many-Task Application	Lee Berry	Oak Ridge National Laboratory
	Best Paper Award Finalist	D. Samaddar	ITER Organization
		David Newman	University of Alaska, Fairbanks
		Raul Sanchez	Universidad Carlos III de Madrid
2:30PM		Khushbu Agarwal	Pacific Northwest National Lab
	Design and Involution at the Develop Table Helmid Cale	Jared chase	Pacific Northwest National Lab
	Design and Implementation of 'Many Parallel Task' Hybrid Sub		Pacific Northwest National Lab
	surface Model	Timothy Scheibe	Pacific Northwest National Lab
		Bruce Palmer	Pacific Northwest National Lab
		Todd Elsethagen	Pacific Northwest National Lab
3:00PM	High Doufesses on Matrix Investige Douglas III Footsvisation	Jack Dongarra	University of Tennessee
	High Performance Matrix Inversion Based on LU Factorization	Mathieu Faverge	ICL
	for Multicore Architectures	Hatem Ltaief	KAUST Supercomputing Lab
		Piotr Luszczek	University of Tennessee
3:30PM	Break		
4:00PM	Toward Scalable I/O Architecture for Exascale Systems	Yong Chen	Texas Tech University
4:30PM	MATE-EC2: A Middleware for Processing Data with AWS	Tekin Bicer	The Ohio State University
		David Chiu	Washington State University
	Invited Paper	Gagan Agrawal	The Ohio State University
5:00PM		Elif Dede	SUNY Binghamton
	Riding the Elephant: Managing Ensembles with Hadoop	Madhusudan Govindaraju	SUNY Binghamton
		Dan Gunter	Lawrence Berkeley National Lab
		Lavanya Ramakrishnan	Lawrence Berkeley National Lab
5:30PM	Best Paper Award		
	Attendee Prize Giveaway Win an Apple iPad 2 Tablet	Ioan Raicu	Illinois Institute of Technology & Argonne National Lab.
		Yong Zhao	University of Electronic Science and Technology of China
	Register at <a href="http://datasys.cs.iit.edu/events/MTAGS11/prize.html">http://datasys.cs.iit.edu/events/MTAGS11/prize.html</a>		

# Organizers



loan Raicu
DataSys Laboratory
Computer Science Department
Illinois Institute of Technology
Math and Computer Science Div.
Argonne National Laboratory



Yong Zhao School of Computer Science and Engineering University of Electronic Science and Technology of China



# lan T. Foster Computer Science Department Computation Institute University of Chicago Math and Computer Science Div. Argonne National Laboratory













Computational and Data Driven Sciences have become the third and fourth pillar of empirical sciences in addition to experimental and theoretical science. Scientific Computing has already begun to change how science is done, enabling scientific breakthroughs through new kinds of experiments that would have been impossible only a decade ago. Today's science is generating datasets that are increasing exponentially in both complexity and volume, making their analysis, archival, and sharing one of the grand challenges of the 21st century. The support for data intensive computing is critical to advance modern science as storage systems have exposed a widening gap between their capacity and their bandwidth by more than 10-fold over the last decade. There is a growing need for advanced techniques to manipulate, visualize and interpret large datasets. Scientific Computing is the key to solving "grand challenges" in many domains and provide breakthrough in new knowledge, and it comes in many shapes and forms, from high-performance computing (HPC) which is heavily focused on compute-intensive applications, high-throughput computing (HTC) which focuses on using many computing resources over long periods of time to accomplish its computational tasks, many-task computing (MTC) which aims to bridge the gap between HPC and HTC by focusing on using many resources over short periods of time, to data-intensive computing which is heavily focused on data distribution and harnessing data locality by scheduling of computations close to the data.

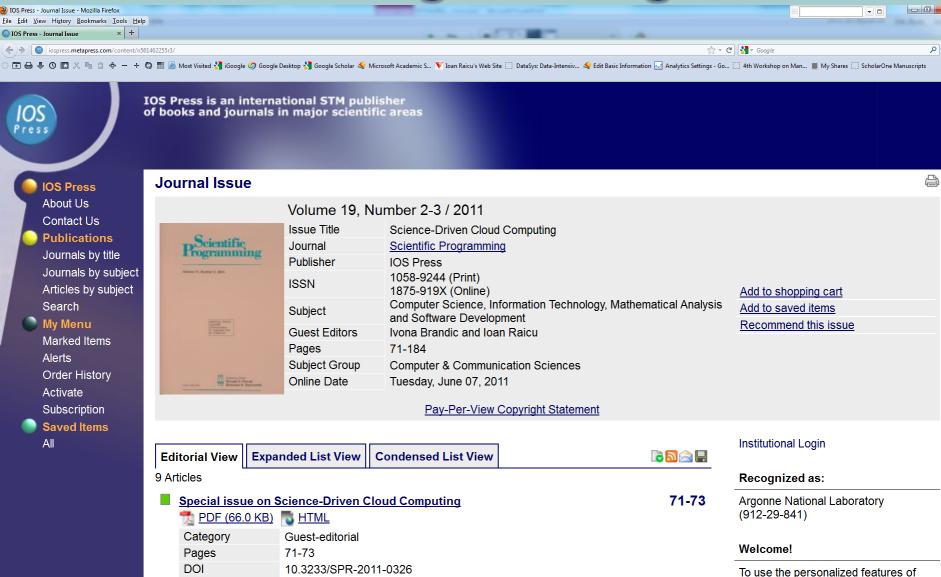
The 3<sup>rd</sup> workshop on Scientific Cloud Computing (ScienceCloud) will provide the scientific community a dedicated forum for discussing new research, development, and deployment efforts in running these kinds of scientific computing workloads on Cloud Computing infrastructures. The ScienceCloud workshop will focus on the use of cloud-based technologies to meet new compute intensive and data intensive scientific challenges that are not well served by the current supercomputers, grids and HPC clusters. The workshop will aim to address questions such as: What architectural changes to the current cloud frameworks (hardware, operating systems, networking and/or programming models) are needed to support science? Dynamic information derived from remote instruments and coupled simulation, and sensor ensembles that stream data for realiment analysis are important emerging techniques in scientific and cyber-physical engineering systems. How can cloud technologies enable and adapt to these new scientific approaches dealing with dynamism? How are scientists using clouds? Are there scientific HPC/HTC/MTC workloads that are suitable candidates to take advantage of emerging cloud computing resources with high efficiency? Commercial public clouds provide easy access to cloud infrastructure for scientists. What are the gaps in commercial cloud offerings and how can they be adapted for running existing and novel eScience applications? What benefits exist by adopting the cloud model, over clusters, grids, or supercomputers? What factors are limiting clouds use or would make them more usable/efficient?

This workshop encourages interaction and cross-pollination between those developing applications, algorithms, software, hardware and networking, emphasizing scientific computing for such cloud platforms. We believe the workshop will be an excellent place to help the community define the current state, determine future goals, and define architectures and services for future science clouds.

# **Topics of Interest**

We invite the submission of original work that is related to the topics below. The papers can be either short (5 pages) position papers, or long (10 pages) research papers. Topics of interest include (in the context of Cloud Computing):

# Scientific Programming Journal

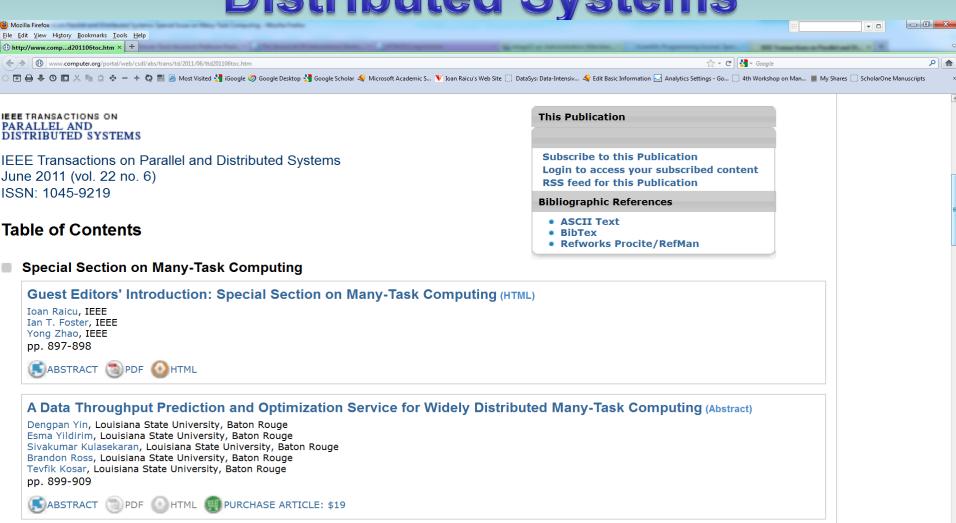


this site, please log in or register.

Ivona Brandic and Ioan Raicu

Authors

# IEEE Transaction on Parallel and Distributed Systems



Abdullah Gharaibeh, The University of British Columbia, Vancouver Samer Al-Kiswany, The University of British Columbia, Vancouver

ThriftStore: Finessing Reliability Trade-Offs in Replicated Storage Systems (Abstract)

# **Springer Journal of Grid Computing**

# Special Issue on Data Intensive Computing in the Clouds

http://datasys.cs.iit.edu/events/JGC-DataCloud-2012/index.html

### **Topics**

Authors are invited to submit original and unpublished technical papers in these topics:

- Data-intensive cloud computing applications, characteristics, challenges
- Case studies of data intensive computing in the clouds
- Performance evaluation of data clouds, data grids, and data centers
- Energy-efficient data cloud design and management
- Data placement, scheduling, and interoperability in the clouds
- Accountability, QoS, and SLAs
- Data privacy and protection in a public cloud environment
- Distributed file systems for clouds
- Data streaming and parallelization
- New programming models for data-intensive cloud computing
- Scalability issues in clouds
- Social computing and massively social gaming intensive computing.
- 3D Internet and implications
- Future research challenges in data-intensive cloud computing

### **Guest Editors**

loan Raicu, Illinois Institute of Technology

& Argonne National Laboratory

Tevfik Kosar, University at Buffalo

### **Editors-in-Chief**

Peter Kacsuk, Hungarian Academy of Sciences lan Foster, University of Chicago & Argonne National Laboratory

### Overview

Applications and experiments in all areas of science are becoming increasingly complex and more demanding in terms of their computational and data requirements. Some applications generate data volumes reaching hundreds of terabytes and even petabytes. As scientific applications become more data intensive, the management of data resources and dataflow between the storage and compute resources is becoming the main bottleneck. Analyzing, visualizing, and disseminating these large data sets has become a major challenge and data intensive computing is now considered as the "fourth paradigm" in scientific discovery after empirical, theoretical, and computational scientific approaches.

The Special Issue on Data Intensive Computing in the Clouds will provide the scientific community a dedicated forum, within the prestigious Springer Journal of Grid Computing, for presenting new research, development, and deployment efforts in running data-intensive computing workloads on Cloud Computing infrastructures. This special issue will focus on the use of cloud-based technologies to meet the new data intensive scientific challenges that are not well served by the current supercomputers, grids or compute-intensive clouds. We believe this venue will be an excellent place to help the community define the current state, determine future goals, and present architectures and services for future clouds supporting data intensive computing.

# **Important Dates**

Papers Due: August 16th, 2011

First Round Decision: October 15th, 2011 Second Round Decision: December 15th, 2011 Final Decision: February 1st, 2012 Publication Date: June 2012

Web Site: http://datasys.cs.iit.edu/events/IGC-DataCloud-2012/ Submission Site: http://grid.edmgr.com/

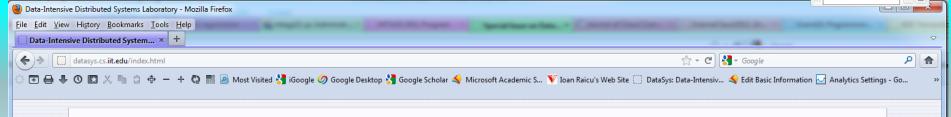
Contact Guest Editors: jgc-datacloud-2012@datasys.cs.iit.edu













# **Data-Intensive Distributed Systems Laboratory**

<u>Illinois Institute of Technology</u> Department of Computer Science

Home

News

**Projects** 

People

**Publications** 

Theses & Dissertations

**Events** 

CS Seminar

DataSys Seminar

Courses

Links

Contact

Mailing List

Wiki

The DataSys Lab is a research lab in the Department of Computer Science at Illinois Institute of Technology, being lead by Dr. Ioan Raicu. The DataSys Lab conducts research in various areas of distributed systems with an emphasis on designing, implementing, and evaluating systems, protocols, and middleware with the goal of supporting data-intensive applications at extreme scales. The lab's mission is to investigate challenging, high-impact research projects to support data-intensive distributed computing on a variety of systems, from many-core systems, clusters, grids, clouds, and supercomputers.

### **News**

- August 21st, 2011: New Software Release: CiteSearcher v1.0
- August 22nd, 2011: New Course: CS553, Cloud Computing
- August 4th, 2011: News Story in HPCinTheCloud: Cloud, Exascale Challenges Converge
- July 17th, 2011: The Second International Workshop on Data Intensive Computing in the Clouds (DataCloud-SC11)

# amh, looking for bright PhD students

# **Active Projects**

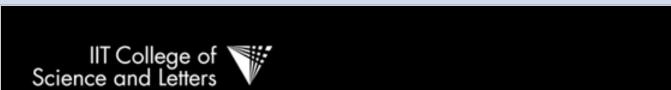
- <u>dFalkon: distributed Fast and Light-weight tasK executiON framework</u>
- FusionFS: Fusion distributed File System
- ZHT: Zero-Hop Distributed Hash Table
- Falkon: Fast and Light-weight tasK executiON framework
- Swift: Fast, Reliable, Loosely Coupled Parallel Computation
- CiteSearcher: a Google Scholar frontend for iOS and Android mobile devices

Home | News | Projects | People | Publications | Theses & Dissertations | Events | CS Seminar | DataSys Seminar | Courses | Links | Contact | Wiki

# New MS Specialization in Distributed and Cloud Computing

File Edit View History Bookmarks Tools Help VIIT Science and Letters | Computer ... × + 

📭 😭 🍖 😑 🕇 🔯 🛗 Most Visited 🛂 iGoogle 🧭 Google Desktop 🛂 Google Desktop 🤻 Google Scholar 🦂 Microsoft Academic S... 🔻 Ioan Raicu's Web Site 💮 DataSys: Data-Intensiv... 🗳 Edit Basic Information 🖂 Analytics Settings - Go... 💮 4th Workshop on Man... 🔳 My Shares 💮 ScholarOne Manuscripts



Academic Programs / Graduate

**Computer Science Graduate Programs** 



ILLINOIS INSTITUTE OF TECHNOLOGY

CSI Home Graduate Admission Undergrad Admission Financial Aid **CSL Programs** Student Life

IIT.EDU HOME

# Apply Now » PROSPECTIVE STUDENTS **ALUMNI BUSINESS & INDUSTRY VISITORS** Quick Links Search Content Contact CSL **COMPUTER SCIENCE**

### COMPUTER SCIENCE HOME

**ABOUT PEOPLE** 

### ACADEMIC PROGRAMS

Undergraduate

Degrees Offered

Course Descriptions

### Graduate

Degrees Offered

Course Descriptions Acceleusted Comment 30 credit hours

Master Of Computer Science With a Specialization in Distributed and Cloud Computing

The Master of Computer Science With a Specialization in Distributed and Cloud Computing is intended for students who are interested to learn about distributed systems and how they are applied to real world problems, as well as how emerging cloud computing technologies can be used to implement some of the worlds most popular services and applications.



# Tenure-Track Assistant Professor Position Available for Fall 2012

# Department of Computer Science Illinois Institute of Technology Chicago, Illinois

Applications are invited for a tenure-track assistant professor position in Computer Science beginning Fall 2012. Excellence in research, teaching and obtaining external funding is expected. While strong candidates from all areas of computer science will be considered, applicants from general data areas such as database, data mining, information security, information retrieval, and data understanding and processing are especially encouraged.

The Department offers B.S., M.S., and Ph.D. degrees in Computer Science and has research strengths in distributed systems, information retrieval, computer networking, intelligent information systems and algorithms. The Illinois Institute of Technology, located within 10 minutes of downtown Chicago, is a dynamic and innovative institution. The Department has strong connections to Argonne and Fermi National Laboratories, and to local industry, and is on a successful and aggressive recruitment plan. IIT is an equal opportunity/affirmative action employer. Women and Underrepresented Minorities are strongly encouraged to apply.

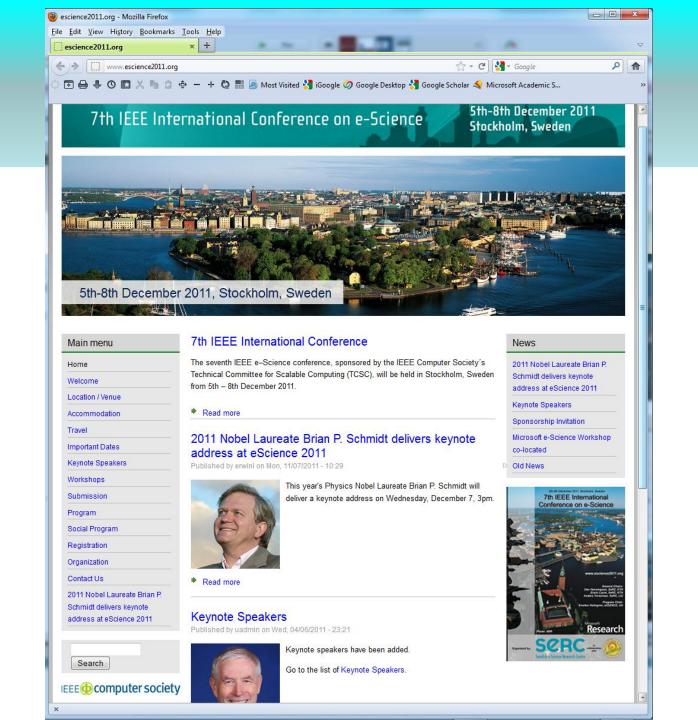
Evaluation of applications will start on December 1, 2011 and will continue until the position is filled. Applicants must complete the online form at <a href="https://docs.google.com/spreadsheet/viewform?formkey=dFRDbTBBVnVmYTFiYk1pblpxYVI1MWc6MQ">https://docs.google.com/spreadsheet/viewform?formkey=dFRDbTBBVnVmYTFiYk1pblpxYVI1MWc6MQ</a> (the online form cannot be updated once submitted, any updates should be sent via email to <a href="mailto-search@cs.iit.edu">search@cs.iit.edu</a> with "updated online form information" in the subject); in addition, applications must submit a detailed curriculum vita, a statement of research and teaching interests, and the names and email addresses of at least four references to:

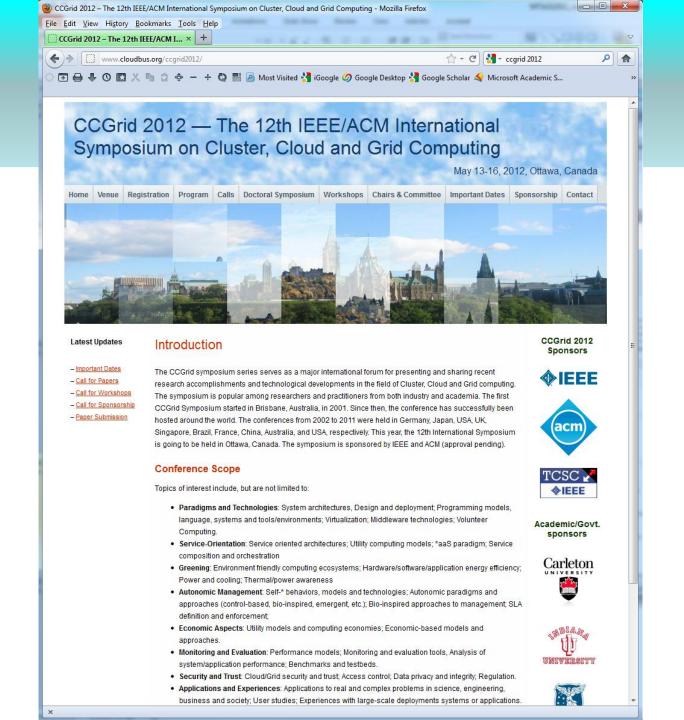
### Computer Science Faculty Search Committee

Department of Computer Science Illinois Institute of Technology

# Focusing on the general data area!

Email: search@cs.iit.edu http://www.iit.edu/csl/cs







### Menu

# Home Important Dates

Conference Poster

# Organization

Papers

Call for Papers
Paper Submission
Camera Ready

## Program

Keynote Speakers Conference Program

Social Events

# Workshops

Call for Workshops

### Venue

General Information

Travel to Delft

Travel in Delft

Accommodation

### Registration

**HPDC Conference Series** 

## ACM Symposium on High-Performance Parallel and Distributed Computing

# Welcome to HPDC'12

The organizing committee is delighted to invite you to **HPDC'12**, the 21st International ACM Symposium on High-Performance Parallel and Distributed Computing, to be held in **Delft, the Netherlands**, which is a historic, picturesque city that is less than one hour away from Amsterdam-Schiphol airport.

HPDC is the premier annual conference on the design, the implementation, the evaluation, and the use of parallel and distributed systems for high-end computing. HPDC is sponsored by SIGARCH, the <a href="Special Interest Group on Computer Architecture">Special Interest Group on Computer Architecture of the Association for Computing Machinery</a>.

HPDC'12 will be held at <u>Delft University of Technology</u>, with the main conference taking place on **June 20-22** (Wednesday to Friday), and with affiliated workshops on **June 18-19** (Monday and Tuesday).

Here is a presentation with a preview of HPDC'12 in Delft.

**Note:** This website is still **under construction** and several pages are still empty. They will be filled in due course.

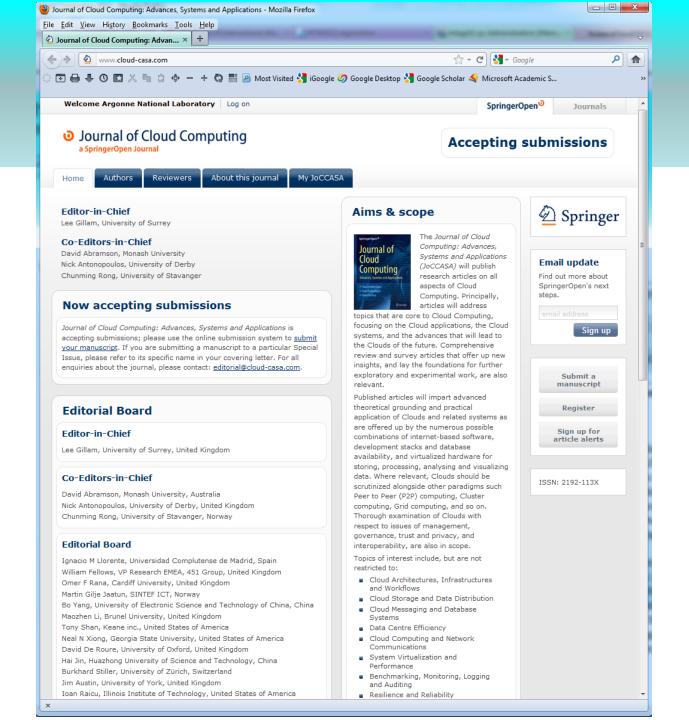
# Supported By



Delft University of Technology



© 2011 | Delft University of Technology | Department of Software Technology | Parallel and Distributed Systems



# More Information

- DataCloud-SC11 Website:
  - http://datasys.cs.iit.edu/events/DataCloud-SC11/
- Workshop program
  - http://datasys.cs.iit.edu/events/DataCloud-SC11/program.html
- Prize giveaway (win an Amazon Kindle Fire):
  - http://datasys.cs.iit.edu/events/DataCloud-SC11/prize.html
- Contact:
  - datacloud-sc11-chairs@datasys.cs.iit.edu