

2015 IEEE/ACM 2nd International Symposium on Big Data Computing (BDC 2015)

**Limassol, Cyprus
7-10 December 2015**



**IEEE Catalog Number: CFP15B46-POD
ISBN: 978-1-5090-0340-2**

**Copyright © 2015, The Association for Computing Machinery (ACM)
All Rights Reserved**

******This publication is a representation of what appears in the IEEE
Digital Libraries. Some format issues inherent in the e-media version may
also appear in this print version.***

IEEE Catalog Number:	CFP15B46-POD
ISBN (Print-On-Demand):	978-1-5090-0340-2
ISBN (Online):	978-0-7695-5696-3

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2015 2nd IEEE/ACM International Symposium on Big Data Computing

BDC 2015

Table of Contents

Message from General and Program Chairs	vii
Symposium Organization.....	ix
Program Committee.....	x

BDC1: Big Data Infrastructure

Alleviation of Disk I/O Contention in Virtualized Settings for Data-Intensive Computing	1
<i>Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara</i>	
FaBRiQ: Leveraging Distributed Hash Tables towards Distributed Publish-Subscribe Message Queues	11
<i>Iman Sadooghi, Ke Wang, Dharmit Patel, Dongfang Zhao, Tonglin Li, Shiva Srivastava, and Ioan Raicu</i>	
Towards a Hybrid Imputation Approach Using Web Tables	21
<i>Ahmad Ahmadov, Maik Thiele, Julian Eberius, Wolfgang Lehner, and Robert Wrembel</i>	

BDC2: Big Data Applications

An Empirical Approach to Detection of Topic Bubbles in Tweets	31
<i>Ashwin Kumar TK, K.M. George, and Johnson P. Thomas</i>	
Building the Dresden Web Table Corpus: A Classification Approach	41
<i>Julian Eberius, Katrin Braunschweig, Markus Hentsch, Maik Thiele, Ahmad Ahmadov, and Wolfgang Lehner</i>	
Big Data Remote Access Interfaces for Light Source Science	51
<i>Justin M. Wozniak, Kyle Chard, Ben Blaiszik, Ray Osborn, Michael Wilde, and Ian Foster</i>	

BDC3: Machine Learning & Data Analytics

Conformal Prediction in Spark: Large-Scale Machine Learning with Confidence	61
<i>Marco Capuccini, Lars Carlsson, Ulf Norinder, and Ola Spjuth</i>	
Shock: Active Storage for Multicloud Streaming Data Analysis	68
<i>Andreas Wilke, Wolfgang Gerlach, Travis Harrison, Tobias Paczian, Wei Tang, William Trimble, Jared Wilkening, Narayan Desai, and Folker Meyer</i>	
A Unified Computation Engine for Big Data Analytics	73
<i>Chenyang Xu, Yanjie Chen, Qin Liu, Weixiong Rao, Hong Min, and Gong Su</i>	

BDC4: Scalability & Middleware

Policy-Driven Data Management Middleware for Multi-cloud Storage in Multi-tenant SaaS	78
<i>Ansar Rafique, Dimitri Van Landuyt, Bert Lagaisse, and Wouter Joosen</i>	
Any Data, Any Time, Anywhere: Global Data Access for Science	85
<i>Kenneth Bloom, Tommaso Boccali, Brian Bockelman, Daniel Bradley, Sridhara Dasu, Jeff Dost, Federica Fanzago, Igor Sfiligoi, Alja Mrak Tadel, Matevz Tadel, Carl Vuosalo, Frank Würthwein, Avi Yagil, and Marian Zvada</i>	
Linked 'Big' Data: Towards a Manifold Increase in Big Data Value and Veracity	92
<i>Jeremy Debattista, Christoph Lange, Simon Scerri, and Sören Auer</i>	
Assessing Process Discovery Scalability in Data Intensive Environments	99
<i>Sergio Hernández, Joaquín Ezpeleta, S.J. van Zelst, and Wil M. P. van der Aalst</i>	
Author Index	105