

**2016 IEEE Second International  
Conference on Big Data  
Computing Service  
and Applications  
(BigDataService 2016)**

**Oxford, United Kingdom  
29 March – 1 April 2016**



**IEEE Catalog Number: CFP16A91-POD  
ISBN: 978-1-5090-2252-6**

**Copyright © 2016 by the Institute of Electrical and Electronics Engineers, Inc  
All Rights Reserved**

*Copyright and Reprint Permissions:* Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. 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:	CFP16A91-POD
ISBN (Print-On-Demand):	978-1-5090-2252-6
ISBN (Online):	978-1-5090-2251-9

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# 2016 IEEE Second International Conference on Big Data Computing Service and Applications

## BigDataService 2016

### Table of Contents

Message from General Chairs.....	X
Message from Program Chairs.....	xi
Conference Organization.....	xii
Program Track Committees.....	xv

---

#### Big Data Computing and Service

Cartesian Operations on Distributed Datasets Using Virtual Partitioning .....	1
<i>Michael Phinney, Sean Lander, Matt Spencer, and Chi-Ren Shyu</i>	
Benchmarking of Distributed Computing Engines Spark and GraphLab for Big Data Analytics .....	10
<i>Jian Wei, Kai Chen, Yi Zhou, Qu Zhou, and Jianhua He</i>	
Spark-Based Political Event Coding .....	14
<i>Mohiuddin Solaimani, Rajeevardhan Gopalan, Latifur Khan, Patrick T. Brandt, and Bhavani Thuraisingham</i>	
A Multi-key Transactions Model for NoSQL Cloud Database Systems .....	24
<i>Adewole Ogunyadeka, Muhammad Younas, Hong Zhu, and Arantza Aldea</i>	

#### Big Data Social Networks

Predicting Gaming Related Properties from Twitter Profiles .....	28
<i>Alfredo Kalaitzis, Maria Ivanova Gorinova, Yoad Lewenberg, Yoram Bachrach, Michael Fagan, Dean Carignan, and Nitin Gautam</i>	
Mining User Interests to Predict Perceived Psycho-Demographic Traits on Twitter .....	36
<i>Svitlana Volkova, Yoram Bachrach, and Benjamin Van Durme</i>	
Using N-Gram Graphs for Sentiment Analysis: An Extended Study on Twitter .....	44
<i>Fotis Aisopos, Dimitrios Tzannetos, John Violos, and Theodora Varvarigou</i>	

Opinion Mining and Sentiment Polarity on Twitter and Correlation between Events and Sentiment .....	52
<i>Peiman Barnaghi, Parsa Ghaffari, and John G. Breslin</i>	

## **Big Data Industry Track**

Attribute Based Access Control for Big Data Applications by Query Modification .....	58
<i>Jim Longstaff and Joanne Noble</i>	
Large-Scale Industrial Alarm Reduction and Critical Events Mining Using Graph Analytics on Spark .....	66
<i>Sam Ade Jacobs and Aldo Dagnino</i>	
Fault-Diagnosis for Reciprocating Compressors Using Big Data .....	72
<i>Keerqinhu, Guanqiu Qi, Wei-Tek Tsai, Yi Hong, Wenxiang Wang, Guangxin Hou, and Zhiqin Zhu</i>	
Fortune, Albeit Digital, at the Bottom of the Pyramid - Big Data Powered Business Model for Internet Service Providers .....	82
<i>Aakash Mangal, Adwait Kaley, Arpit Patel, Chandrasekar Vuppalapati, Saumeel Gajera, and Shivang Doshi</i>	

## **Big Data Intelligence and Knowledge Engineering**

TopicListener: Observing Key Topics from Multi-channel Speech Audio Streams .....	85
<i>Jing Su and Oisín Boydell</i>	
Fuzzy Based Clustering Algorithms to Handle Big Data with Implementation on Apache Spark .....	95
<i>Neha Bharill, Aruna Tiwari, and Aayushi Malviya</i>	
A Practical Approach to Hard Disk Failure Prediction in Cloud Platforms: Big Data Model for Failure Management in Datacenters .....	105
<i>Sandipan Ganguly, Ashish Consul, Ali Khan, Brian Bussone, Jacqueline Richards, and Alejandro Miguel</i>	
An Iterative Graph-Based Generic Single and Multi Document Summarization Approach Using Semantic Role Labeling and Wikipedia Concepts .....	117
<i>Muhidin Mohamed and Mourad Oussalah</i>	

## **Big Data Applications in Transportation, Tourism and Smart City**

A Novel Point of Interest (POI) Location Based Recommender System Utilizing User Location and Web Interactions .....	121
<i>Mayy Habayeb, Behjat Soltanifar, Bora Caglayan, and Ayse Bener</i>	
Big Data from Cellular Networks: Real Mobility Scenarios for Future Smart Cities .....	131
<i>Davide Tosi and Stefano Marzorati</i>	
A Hybrid Geometric Approach for Measuring Similarity Level Among Documents and Document Clustering .....	142
<i>Arash Heidarian and Michael J. Dinneen</i>	
Big Data Analytics on Aviation Social Media: The Case of China Southern Airlines on Sina Weibo .....	152
<i>Sien Chen, Yinghua Huang, and Wengqiang Huang</i>	

## **Big Data Application Tracks of Energy and Environment**

The Matsu Wheel: A Cloud-Based Framework for Efficient Analysis and Reanalysis of Earth Satellite Imagery .....	156
<i>Maria T. Patterson, Nikolas Anderson, Collin Bennett, Jacob Bruggemann, Robert L. Grossman, Matthew Handy, Vuong Ly, Daniel J. Mandl, Shane Pederson, James Pivarski, Ray Powell, Jonathan Spring, Walt Wells, and John Xia</i>	
Temporal Uncertainty of Wind Ramp Predictions Using Probabilistic Forecasting Technique .....	166
<i>Yaqiong Li, Petr Musilek, Edward Lozowski, Chaobo Dai, Tongxun Wang, and Zhengang Lu</i>	
Cleaning Framework for BigData: An Interactive Approach for Data Cleaning .....	174
<i>Hong Liu, Ashwin Kumar Tk, Johnson P Thomas, and Xiaofei Hou</i>	
Air Quality Simulations Using Big Data Programming Models .....	182
<i>Haripriya Ayyalasomayajula, Edgar Gabriel, Peggy Lindner, and Daniel Price</i>	

## **Main Track I**

A Versatile Platform for Instrumentation of Knowledge Worker's Computers to Improve Information Analysis .....	185
<i>Paul Jones, Sidharth Thakur, Steven Cox, and Michael Matthews</i>	
Efficient Cloud-Based Framework for Big Data Classification .....	195
<i>Rezvan Pakdel and John Herbert</i>	
A User-Centric Approach to Pricing Information .....	202
<i>Divya Rao and Wee Keong Ng</i>	

## **Main Track II**

Big Data Quality: A Roadmap for Open Data .....	210
<i>Paolo Ciancarini, Francesco Poggi, and Daniel Russo</i>	
Mining and Visualizing Associations of Concepts on a Large-Scale Unstructured Data .....	216
<i>Reza Sadoddin and Osvaldo Driollet</i>	
Policy-Based QoS Enforcement for Adaptive Big Data Distribution on the Cloud .....	225
<i>Hadeel T. El Kassabi, Ikbai Taleb, Mohamed Adel Serhani, and Rachida Dssouli</i>	
Individual Differences and Mobile Service Adoption: An Empirical Analysis .....	234
<i>Runhua Xu, Remo Manuel Frey, and Alexander Ilic</i>	

## **Big Data Security**

A Confidentiality-Preserving Search Technique for Encrypted Relational Cloud Databases .....	244
<i>Vasily Sidorov and Wee Keong Ng</i>	
Multiple Rounds Aware Power Analysis Attack for a Lightweight Cipher SIMECK .....	252
<i>Masaya Yoshikawa, Yusuke Nozaki, and Kensaku Asahi</i>	
Real-Time Discovery Services over Large, Heterogeneous and Complex Healthcare Datasets Using Schema-Less, Column-Oriented Methods .....	257
<i>Edmon Begoli, Ted Dunning, and Charlie Frasure</i>	
Long-Term Monitoring of Wind Farms Using Big Data Approach .....	265
<i>Jan Helsen, Gert De Sitter, and Pieter Jan Jordaens</i>	

## **Big Data Bio-medical and Healthcare Services**

SDenPeak: Semi-supervised Nonlinear Clustering Based on Density and Distance .....	269
<i>Wen-Qi Fan, Chang-Dong Wang, and Jian-Huang Lai</i>	
Dependency Networks Extractions from Textual Sources in Criminology: An Initial Implementation .....	276
<i>Marcello Trovati, Philip Hodgson, Charlotte Hargreaves, Andrew David Baker, Lu Liu, and Nik Bessis</i>	
Building a Core Biomedical Knowledge Base .....	283
<i>Andleeb Shahnaz, Usman Qamar, and Abdul Wahab Muzaffar</i>	
Weighted Multi-view On-Line Competitive Clustering .....	286
<i>Guang-Yu Zhang, Dong Huang, Chang-Dong Wang, and Wei-Shi Zheng</i>	

The Role of Big Data in Creating Sense EHR, an Integrated Approach to Create Next Generation Mobile Sensor and Wearable Data Driven Electronic Health Record (EHR) .....	293
<i>Chandrasekar Vuppalapati, Anitha Ilapakurti, and Santosh Kedari</i>	
<b>Author Index</b> .....	297