

**2017 IEEE Third International
Conference on Big Data
Computing Service and
Applications
(BigDataService 2017)**

**Redwood City, California, USA
6-9 April 2017**



**IEEE Catalog Number: CFP17A91-POD
ISBN: 978-1-5090-6319-2**

**Copyright © 2017 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 is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP17A91-POD
ISBN (Print-On-Demand):	978-1-5090-6319-2
ISBN (Online):	978-1-5090-6318-5

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

2017 IEEE Third International Conference on Big Data Computing Service and Applications

BigDataService 2017

Table of Contents

Message from Big Data Service 2017 General Chairs	x
Message from Big Data Service 2017 Program Chairs	xi
Big Data Service 2017 Organizing Committee	xii
Big Data Service 2017 Program Committee	xiii
Smart City Big Data Workshop Chairs and Committee	xv
Big Data Security and Service Workshop Chairs and Committee	xvi
Big Data Quality Assurance Workshop Chairs and Committee	xvii
Keynotes	xviii

Session 1: Intelligent Data Mining

Enriched Over_Sampling Techniques for Improving Classification of Imbalanced Big Data	1
<i>Sachin Subhash Patil and Shefali Pratap Sonavane</i>	
Improved Sentiment Classification by Multi-Modal Fusion	11
<i>Lige Gan, Rachid Benlamri, and Richard Khoury</i>	
Towards Automatic Linkage of Knowledge Worker's Claims with Associated Evidence from Screenshots	17
<i>Paul Jones, Dakota Medd, Sreekanth Ramakrishnan, Rajat Shah, Joann Keyton, and Nagiza Samatova</i>	

Session 2: Frameworks for Big Data Processing

CaPaR: A Career Path Recommendation Framework	23
<i>Bharat Patel, Varun Kakuste, and Magdalini Eirinaki</i>	
A Goal-Oriented Big Data Analytics Framework for Aligning with Business	31
<i>Grace Park, Lawrence Chung, Liping Zhao, and Sam Supakkul</i>	

When Rule Engine Meets Big Data: Design and Implementation of a Distributed Rule Engine Using Spark	41
<i>Jindou Zhang, Jinxing Yang, and Jing Li</i>	

Session 3: Big Data Processing and Mining

Balanced Parallel Frequent Pattern Mining over Massive Data Stream	50
<i>Xi Fu, Lei Shi, and Jing Li</i>	
Data Allocation of Large-Scale Key-Value Store System Using Kinetic Drives	60
<i>Xiang Cao, Manas Minglani, and David Hung-Chang Du</i>	
Scaling Collaborative Filtering to Large-Scale Bipartite Rating Graphs Using Lenskit and Spark	70
<i>Christos Sardanios, Iraklis Varlamis, and Magdalini Eirinaki</i>	
Small Boxes Big Data: A Deep Learning Approach to Optimize Variable Sized Bin Packing	80
<i>Feng Mao, Edgar Blanco, Mingang Fu, Rohit Jain, Anurag Gupta, Sebastien Mancel, Rong Yuan, Stephen Guo, Sai Kumar, and Yayang Tian</i>	

Session 4: Big Data Analysis

Impact of Core Charging Station's Cease Operation in the Entire Charging Station System: A Case Study in Shenzhen	90
<i>Zhiyong Tian, Lai Tu, Yi Wang, Fan Zhang, and Chen Tian</i>	
MD-Miner: Behavior-Based Tracking of Network Traffic for Malware-Control Domain Detection	96
<i>Jia-Hao Sun, Tzung-Han Jeng, Chien-Chih Chen, Hsiu-Chuan Huang, and Kuo-Sen Chou</i>	

Smart City Big Data Workshop

Session 1: Smart City Big Data

Viral Marketing for Smart Cities: Influencers in Social Network Communities	106
<i>Madhura Kaple, Ketki Kulkarni, and Katerina Potika</i>	
A Deep Learning Framework for Smart Street Cleaning	112
<i>Chandni Balchandani, Rakshith Koravadi Hatwar, Parteek Makkar, Yanki Shah, Pooja Yelure, and Magdalini Eirinaki</i>	
Smart Dairies — Enablement of Smart City at Gross Root Level	118
<i>Jaya Shankar Vuppalapati, Santosh Kedari, Ananth Ilapakurthy, Anitha Ilapakurti, and Chandrasekar Vuppalapati</i>	
iDispenser — Big Data Enabled Intelligent Dispenser	124
<i>Anitha Ilapakurti, Jaya Shankar Vuppalapati, Santosh Kedari, Sharat Kedari, Chitanshu Chauhan, and Chandrasekar Vuppalapati</i>	

Session 5: Data Analytics and Visualization

Engineering Scalable Distributed Services for Real-Time Big Data Analytics	131
<i>Sahar Jambi and Kenneth M. Anderson</i>	
Answering How-to-Reach Query in Big Attributed Graph Databases	141
<i>Duncan Yung and Shi-Kuo Chang</i>	
On Building a Big Data Analysis System for California Drought	149
<i>Pengcheng Zhang, Jerry Gao, A. G. Thomas, K. P. Alagupackiam, K. Mannava, P. I. Bosco, and Sen Chiao</i>	
Rapidly Generate and Visualize the Digest of Massive Time Series Data	157
<i>Zhiwei Gao, Tiangang Zou, Lulu Zhang, Juan Chen, Jing Guo, and Weixing Ji</i>	
Visualization Techniques for Electrical Grid Smart Metering Data: A Survey	165
<i>Maria Stefan, Jose G. Lopez, Morten H. Andreassen, and Rasmus L. Olsen</i>	

Session 6: Big Data Application

Data Reduction as a Service in Smart City Architecture	172
<i>Aseel Alkhalaiwi and Dan Grigoras</i>	
Wind Farm Operation and Maintenance Optimization Using Big Data	179
<i>Jan Helsen, Cédric Peeters, Peter Doro, Eveline Ververs, and Pieter Jan Jordaens</i>	
Automated Medical Diagnosis from Clinical Data	185
<i>Vishnu S. Pendyala and Silvia Figueira</i>	

Session 7: Big Data Framework, Technology, and Solutions

A Distributed Multi-exemplar Affinity Propagation Clustering Algorithm Based on MapReduce	191
<i>Yu-Bo Yang, Chang-Dong Wang, and Jian-Huang Lai</i>	
CENSE: A Cognitive Navigation System for People with Special Needs	198
<i>Akhila Kishore, Anhad Bhasin, Arun Balaji, Chandrasekar Vuppapapati, Divyesh Jadav, Preethi Anantharaman, and Shrutee Gangras</i>	
Transparent Link: A Framework of Anonymizing MSA-Dataset Based on Probabilistic Graphical Model	204
<i>Li-E Wang, Yameng Guo, and Xianxian Li</i>	

Session 8: Invited Papers

An Advanced Inventory Data Mining System for Business Intelligence	210
<i>Qifeng Zhou, Bin Xia, Wei Xue, Chunqiu Zeng, Ruyuan Han, and Tao Li</i>	
A Review of Performance Evaluation on 2D Face Databases	218
<i>Gabriel Castaneda and Taghi M. Khoshgoftaar</i>	
Data-Driven Water Quality Analysis and Prediction: A Survey	224
<i>Gaganjot Kang, Jerry Zeyu Gao, and Gang Xie</i>	

PROV-TE: A Provenance-Driven Diagnostic Framework for Task Eviction in Data Centers	233
<i>Abdulaziz Albatli, David McKee, Paul Townend, Lydia Lau, and Jie Xu</i>	

Big Data Security and Service Workshop

Session 1: Security Services for Smart City

Smart City Service Acceleration on FPGAs	243
<i>Youngsoo Kim, Marcus Garcia, Alan Chen, Nathan Wong, Shrikant Jadhav, and Clay S. Gloster</i>	
Analyzing Movie Scripts as Unstructured Text	249
<i>Seong-Ho Lee, Hye-Yeon Yu, and Yun-Gyung Cheong</i>	
Smart Illegal Dumping Detection	255
<i>Akshay Dabholkar, Bhushan Muthiyar, Shilpa Srinivasan, Swetha Ravi, Hyeran Jeon, and Jerry Gao</i>	
Intelligent Learning Systems Design for Self-Defense Education	261
<i>Hyeran Jeon, Kaikai Liu, Younghee Park, Jerry Gao, Gong Chen, and Jim Kao</i>	

Session 2: Big Data for Security

Big Data Cloud Deduplication Based on Verifiable Hash Convergent Group Signcryption	265
<i>Ei Mon Cho and Takeshi Koshiba</i>	
DDoS Attack Mitigation in Internet of Things Using Software Defined Networking	271
<i>M. Ejaz Ahmed and Hyoungshick Kim</i>	
Graph-Theory Based Simplification Techniques for Efficient Biological Network Analysis	277
<i>Euseong Ko, Mingon Kang, Hyung Jae Chang, and Donghyun Kim</i>	

Big Data Quality Assurance Workshop

Session 1: Big Data Quality Assurance and Validation

Big Data Validation Case Study	281
<i>Chunli Xie, Jerry Gao, and Chuanqi Tao</i>	
A Practical Approach to Validation of Buildings' Sensor Data: A Commissioning Experience Report	287
<i>Claudio Giovanni Mattera, Sanja Lazarova-Molnar, Hamid Reza Shaker, and Bo Nørregaard Jørgensen</i>	
Evolution Slicing-Based Change Impact Analysis	293
<i>Wanzhi Wen, Jianping Chen, Jiaqi Yuan, and Xiaoyong Chen</i>	

Session 2: Quality Assurance and Validation for Big Data-Based Applications

A Quality Evaluation Approach to Search Engines of Shopping Platforms	299
<i>Chen Hao, Tao Chuanqi, and Jerry Gao</i>	
A Viewpoint on Location-Based Mobile Apps Testing	305
<i>Wajeh Addin Mohsen, Chuanqi Tao, Qian-mu Li, and Jerry Gao</i>	
A Survey on Quality Assurance Techniques for Big Data Applications	313
<i>Pengcheng Zhang, Xuewu Zhou, Wenrui Li, and Jerry Gao</i>	
Author Index	320