

2015 IEEE First International Conference on Big Data Computing Service and Applications

(BigDataService 2015)

**Redwood City, California, USA
30 March – 2 April 2015**



**IEEE Catalog Number: CFP15A91-POD
ISBN: 978-1-4799-8129-8**

2015 IEEE First International Conference on Big Data Computing Service and Applications

BigDataService 2015

Table of Contents

Message from the IEEE BigDataService 2015 General	
Chairs	xi
Message from the IEEE BigDataService 2015 Program	
Chairs	xii
IEEE BigDataService 2015 Organizing Committee	xiii
IEEE BigDataService 2015 Program Committee	xvi
Sponsors	xxii

Session 1A: Big Data Networking

Interest-Based Mining and Modeling of Big Mobile Networks	1
<i>Saeed Moghaddam and Ahmed Helmy</i>	
A Hybrid Network Architecture for Data Centers	7
<i>Mohammad Naimur Rahman and Amir Esmailpour</i>	

Session 2A: Big Data Computing and Service (I)

Scalable Security Event Aggregation for Situation Analysis	14
<i>Jinoh Kim, Ilhwan Moon, Kyungil Lee, Sang C. Suh, and Ikkyun Kim</i>	
Big Graph Processing Systems: State-of-the-Art and Open Challenges	24
<i>Radwa Elshawi, Omar Batarfi, Ayman Fayoumi, Ahmed Barnawi, and Sherif Sakr</i>	
Fast Quantitative Analysis of Stock Trading Points in Dual Period of DMAC	34
<i>Yi Wei and Vipin Chaudhary</i>	
An Optimized Generic Client Service API for Managing Large Datasets within a Data Repository	44
<i>Ajinkya Prabhune, Rainer Stotzka, Thomas Jejkal, Volker Hartmann, Margund Bach, Eberhard Schmitt, Michael Hausmann, and Juergen Hesser</i>	

Session 2B: Big Data Sensing and IoT

Monitoring Noise Pollution Using the Urban Civics Middleware	52
<i>Sara Hachem, Vivien Mallet, Raphaël Ventura, Animesh Pathak, Valérie Issarny, Pierre-Guillaume Raverdy, and Rajiv Bhatia</i>	
Performance Characterization of the Servioticy API: An IoT-as-a-Service Data Management Platform	62
<i>Juan Luis Pérez and David Carrera</i>	
Framework Using Bayesian Belief Networks for Utility Effective Management and Operations	72
<i>Joseph Siryani, Thomas Mazzuchi, and Shahram Sarkani</i>	
Big Data Sensing and Service: A Tutorial	79
<i>Jerry Gao, Lihui Lei, and Shui Yu</i>	

Session 3A: Big Data Computing and Service (II)

When Good-Enough is Enough: Complex Queries at Fixed Cost	89
<i>Nathan D. Mickulicz, Rolando Martins, Priya Narasimhan, and Rajeev Gandhi</i>	
Web Service-Enabled Spam Filtering with Naïve Bayes Classification	99
<i>Wanqing You, Kai Qian, Dan Lo, Prabir Bhattacharya, Minzhe Guo, and Ying Qian</i>	
BDAP: A Big Data Placement Strategy for Cloud-Based Scientific Workflows	105
<i>Mahdi Ebrahimi, Aravind Mohan, Andrey Kashlev, and Shiyong Lu</i>	
Do You Feel the Lag of Your Hadoop?	115
<i>Yuxuan Jiang, Zhe Huang, and Danny H.K. Tsang</i>	

Session 3B: Big Data Security

Secure Scalar Product for Big-Data in MapReduce	120
<i>Fang Liu, Wee Keong Ng, and Wei Zhang</i>	
A Sticky Policy Framework for Big Data Security	130
<i>Shuyu Li, Tao Zhang, Jerry Gao, and Younghee Park</i>	
A Secure Video Deduplication Scheme in Cloud Storage Environments Using H.264 Compression	138
<i>Fatema Rashid, Ali Miri, and Isaac Woungang</i>	
Security Implications of Direct-to-Consumer Genetic Services	147
<i>Jingquan Li</i>	

Session 4A: Big Data Computing and Service (III)

Dynamically Scaling Apache Storm for the Analysis of Streaming Data	154
<i>Jan Sipke van der Veen, Bram van der Waaij, Elena Lazovik, Wilco Wijbrandi, and Robert J. Meijer</i>	
Power Efficient MapReduce Workload Acceleration Using Integrated-GPU	162
<i>SungYe Kim, Jeremy Bottleson, Jingyi Jin, Preeti Bindu, Snehal C. Sakhare, and Joseph S. Spisak</i>	
Incremental Sorting for Large Dynamic Data Sets	170
<i>Ahmet Arif Aydin and Kenneth M. Anderson</i>	

Modeling and Designing Fault-Tolerance Mechanisms for MPI-Based MapReduce Data Computing Framework	176
<i>Jian Lin, Fan Liang, Xiaoyi Lu, Li Zha, and Zhiwei Xu</i>	

Session 4B: Big Data Energy and Environment

Predictive Analytics for Extreme Events in Big Data	184
<i>Saahil Shenoy and Dimitry Gorinevsky</i>	
Scalable Geospatial Web Services through Efficient, Online and Near Real-Time Processing of Earth Observation Data	194
<i>Athanasios Karmas, Angelos Tzotsos, and Konstantinos Karantzalos</i>	
Current Developments in Big Data and Sustainability Sciences in Mobile Citizen Science Applications	202
<i>Nikita S. Panchariya, Andrew J. DeStefano, Varsha Nimbagal, Revathi Ragupathy, Serkan Yavuz, Katherine G. Herbert, Emily Hill, and Jerry Alan Fails</i>	
An Efficient Electricity Generation Forecasting System Using Artificial Neural Network Approach with Big Data	213
<i>Mohammad Naimur Rahman and Amir Esmailpour</i>	
The Management of Application of Big Data in Internet of Thing in Environmental Protection in China	218
<i>Sheng Qiang Lu, Gang Xie, Zehua Chen, and Xiaoming Han</i>	

Session 5A: Big Data Modeling, Sharing Technologies and Platforms

AntsBOA: A New Time Series Pipeline for Big Data Processing, Analyzing and Querying in Online Advertising Application	223
<i>Bin Song, Shaosu Liu, Santanu Kolay, and Lawrence Lo</i>	
Towards a 'Big' Health Data Analytics Platform	233
<i>Sangwhan Cha, Ashraf Abusharekh, and Syed SR Abidi</i>	
Industrial Analytics Pipelines	242
<i>Karl Eric Harper, Jiang Zheng, Sam Ade Jacobs, Aldo Dagnino, Anton Jansen, Thomas Goldschmidt, and Adamantios Marinakis</i>	
From Google File System to Omega: A Decade of Advancement in Big Data Management at Google	249
<i>Jade Yang</i>	

Session 5B: Big Data Application and Industry Experience (I)

Building a Big Data Analytics Service Framework for Mobile Advertising and Marketing	256
<i>Lei Deng, Jerry Gao, and Chandrasekar Vuppapalapati</i>	
Favored Encryption Techniques for Cloud Storage	267
<i>Chungsik Song, Younghee Park, Jerry Gao, Sri Kinnera Nanduri, and William Zegers</i>	
Building an IoT Framework for Connected Dairy	275
<i>Anitha Ilapakurti and Chandrasekar Vuppapalapati</i>	

Carotene: A Job Title Classification System for the Online Recruitment Domain	286
<i>Faizan Javed, Qinlong Luo, Matt McNair, Ferosh Jacob, Meng Zhao, and Tae Seung Kang</i>	

Session 6A: Big Data Modeling, Sharing Technologies and Platforms (II)

Table2Graph: A Scalable Graph Construction from Relational Tables Using Map-Reduce	294
<i>Sangkeun Lee, Byung H. Park, Seung-Hwan Lim, and Mallikarjun Shankar</i>	
Index Selection on MapReduce Relational-Databases	302
<i>Fatimah Alsayoud and Ali Miri</i>	
A Low Rank Gaussian Process Prediction Model for Very Large Datasets	308
<i>Roberto Rivera</i>	

Session 6B: Big Data Transportation and Tourism

The Impact of Perceived Security on Consumer E-Loyalty: A Study of Online Tourism Purchasing	314
<i>Feng Cui, Derong Lin, and Yinghua Huang</i>	
Understanding Airline Passenger Behavior through PNR, SOW and Webtrends Data Analysis	323
<i>Sien Chen, Jianping Zhu, Qichang Xie, Wenqiang Huang, and Yinghua Huang</i>	
The 5I Model of Smart City: A Case of Shanghai, china	329
<i>Xixi Lin, Hua Quan, Hong Zhang, and Yinghua Huang</i>	

Session 7A: Big Data Center and Information Banking

A Framework for Multitasking Data-Intensive Management Services in High Performance Computing Environments	333
<i>Sivakumar Kulasekaran, Maria Esteva, Jessica Trelogan, and Si Liu</i>	
Inferring Risk Aversion for Decentralized Investment Portfolios	341
<i>Jason West</i>	
Bankruptcy Prediction of Construction Businesses: Towards a Big Data Analytics Approach	347
<i>Alaka Hafiz, Oyedele Lukumon, Bilal Muhammad, Akinade Olugbenga, Owolabi Hakeem, and Ajayi Saheed</i>	

Session 7B: Big Data Center and Information Banking (II)

Topology-aware Heuristic Data Allocation Algorithm for Big Data Infrastructure	353
<i>Wuhui Chen, Banage T.G.S. Kumara, Incheon Paik, and Zhenni Li</i>	
Understanding Library User Engagement Strategies through Large-Scale Twitter Analysis	361
<i>Hongbo Zou, Hsuanwei Michelle Chen, and Sharmistha Dey</i>	
Semi-supervised Learning on Cross-Lingual Sentiment Analysis with Space Transfer	371
<i>Hui Zhang, Wenhan Chao, and Deqing Wang</i>	
Modeling Both Coarse-Grained and Fine-Grained Topics in Massive Text Data	378
<i>Weifan Zhang, Hui Zhang, Yuan Zuo, and Deqing Wang</i>	

Session 8A: Big Data Intelligence and Knowledge Engineering (I)

Distributed Data Collection, Correlation and Data Reduction for Iterative Streaming Analytics in Wireless Mobile Networks	384
<i>Charles W. Boyle and Surya Kumar Kovvali</i>	
An Evolutionary Rule Mining Method for Continuous Value Prediction from Incomplete Database and Its Application Utilizing Artificial Missing Values	392
<i>Kaoru Shimada, Takaaki Arahira, and Takashi Hanioka</i>	
Data Stream Classification for Structural Health Monitoring via On-Line Support Vector Machines	400
<i>Xiaoou Li and Wen Yu</i>	

Session 8B: Big Data Transportation and Tourism (II)

Research on LBS Impact on College Students' Travel Experience	406
<i>Tu Tingting, Li Yunpeng, Yu Yongyang, Zongqing Zhou, and Yang Xiaohan</i>	
A Study of the Commercial Application of Big Data of the International Hotel Group in China: Based on the Case Study of Marriott International	412
<i>Ying Zhang, Sheng Shu, Zhongjuan Ji, and Yaxin Wang</i>	

Session 9A: Big Data Intelligence and Knowledge Engineering (II)

A Scalable Hierarchical Clustering Algorithm Using Spark	418
<i>Chen Jin, Ruoqian Liu, Zhengzhang Chen, William Hendrix, Ankit Agrawal, and Alok Choudhary</i>	
The Use of Domain Knowledge Models for Effective Data Mining of Unstructured Customer Service Data in Engineering Applications	427
<i>T. Munger, Subhas Desa, and Chris Wong</i>	
An Evolutionary Algorithm for Column Family Schema Optimization in HBase	439
<i>Fangzhou Yang, Dragan Milosevic, and Jian Cao</i>	
Analytics: Key to Go from Generating Big Data to Deriving Business Value	446
<i>Deepali Arora and Piyush Malik</i>	

Session 10A: Bio-Medical and Healthcare Services

MOOC for Medical Big Data Research: An Important Role in Hypertension Big Data Research	453
<i>Xinyan Wang, Liyuan Tian, Bo Xu, Xueliang Wang, and Wenjun Wu</i>	
Bio-medical Application on Predicting Systolic Blood Pressure Using Neural Networks	456
<i>Tony Hao Wu, Enid Wai-Yung Kwong, and Grantham Kwok-Hung Pang</i>	

Session 10B: Big Data Education and Learning

Interactive Visual Text Analysis for Corpus-Based Language Learning	462
<i>Ying Zhu and Eric Friginal</i>	
Philosophy of Big Data: Expanding the Human-Data Relation with Big Data Science Services	468
<i>Melanie Swan</i>	

Session 11A: Multi-media and Social Networking

Finding and Analyzing Principal Features for Measuring User Influence on Twitter	478
<i>Yan Mei, Youliang Zhong, and Jian Yang</i>	
Community-Aware Prediction of Virality Timing Using Big Data of Social Cascades	487
<i>Alvin Junus, Cheung Ming, James She, and Zhanming Jie</i>	

Session 11B: Big Data Application and Industry Experience (II)

WebScalding: A Framework for Big Data Web Services	493
<i>Ferosh Jacob, Aaron Johnson, Faizan Javed, Meng Zhao, and Matt McNair</i>	
Hadoop as Big Data Operating System—The Emerging Approach for Managing Challenges of Enterprise Big Data Platform	499
<i>Sourav Mazumdar and Subhankar Dhar</i>	
Author Index	506