

**2016 IEEE/ACM 3rd
International Conference on Big
Data Computing Applications
and Technologies (BDCAT 2016)**

**Shanghai, China
6-9 December 2016**



**IEEE Catalog Number: CFP16B46-POD
ISBN: 978-1-5090-4468-9**

**Copyright © 2016, Association for Computing Machinery (ACM)
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:	CFP16B46-POD
ISBN (Print-On-Demand):	978-1-5090-4468-9
ISBN (Online):	978-1-4503-4617-7

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

2016 IEEE/ACM 3rd International Conference on Big Data Computing, Applications and Technologies

BDCAT 2016

Table of Contents

Message from the Program Co-Chairs	x
Organizing Committee.....	xi
Program Committee.....	xii

Session 1: Big Data & Machine Learning

Deep Kernel: Learning Kernel Function from Data Using Deep Neural Network	1
<i>Linh Le, Jie Hao, Ying Xie, and Jennifer Priestley</i>	
Empirical Analysis of Asymptotic Ensemble Learning for Big Data	8
<i>Salman Salloum, Joshua Zhexue Huang, and Yulin He</i>	
Spatial Frequency Based Video Stream Analysis for Object Classification and Recognition in Clouds	18
<i>Muhammad Usman Yaseen, Ashiq Anjum, and Nick Antonopoulos</i>	

Session 2: Hadoop & Spark

H2F: A Hierarchical Hadoop Framework for Big Data Processing in Geo-Distributed Environments	27
<i>Marco Cavallo, Giuseppe Di Modica, Carmelo Polito, and Orazio Tomarchio</i>	
Performance Characterization of Hadoop Workloads on SR-IOV-Enabled Virtualized InfiniBand Clusters	36
<i>Shashank Gugnani, Xiaoyi Lu, and Dhabaleswar K. (Dk) Panda</i>	
DRESS: A Rule Engine on Spark for Event Stream Processing	46
<i>Yi Chen and Behzad Bordbar</i>	

Session 3: Visualization & Social Networks

A Visual Analytics Approach to Author Name Disambiguation	52
<i>Chris W. Muelder, Robert Faris, and Kwan-Liu Ma</i>	
Visualization of Big High Dimensional Data in a Three Dimensional Space	61
<i>Ying Xie, Pooja Chenna, Jing (Selena) He, Linh Le, and Jacey Planteen</i>	
Applying Big Data Warehousing and Visualization Techniques on PingER Data	N/A
<i>Aqsa Hameed, Saqib Ali, Rodger Les Cottrell, and Bebo White</i>	
Efficient Service Discovery in Decentralized Online Social Networks	73
<i>Bo Yuan, Lu Liu, and Nick Antonopoulos</i>	

Session 4: Health Applications

Towards Longitudinal Analysis of a Population's Electronic Health Records Using Factor Graphs	79
<i>Arjun P. Athreya, Kee Yuan Ngiam, Zhaojing Luo, E Shyong Tai, Zbigniew Kalbarczyk, and Ravishankar K. Iyer</i>	
Disease Gene Discovery of Single-Gene Disorders Based on Complex Network	87
<i>Chunxiao Xing, Fengjing Shao, Shunyao Wu, and Rencheng Sun</i>	
Identifying Patient Experience from Online Resources via Sentiment Analysis and Topic Modelling	N/A
<i>Mohammed Bahja and Mark Lycett</i>	

Session 5: Data Model & Information Retrieval

Towards a Comprehensive Data LifeCycle Model for Big Data Environments	100
<i>Amir Sinaeepourfard, Jordi Garcia, Xavier Masip-Bruin, and Eva Marín-Tordera</i>	
A Study of Factuality, Objectivity and Relevance: Three Desiderata in Large-Scale Information Retrieval	N/A
<i>Christina Lioma, Birger Larsen, Wei Lu, and Yong Huang</i>	
On Exploiting Data Locality for Iterative MapReduce Applications in Hybrid Clouds	118
<i>Francisco J. Clemente-Castelló, Bogdan Nicolae, Rafael Mayo, Juan Carlos Fernández, and M. Mustafa Rafique</i>	

Session 6: Spatial Data Analytics

CoS-HDFS: Co-Locating Geo-Distributed Spatial Data in Hadoop Distributed File System	123
<i>Mariam Malak Fahmy, Iman Elghandour, and Magdy Nagi</i>	
Spatial and Temporal Analysis of Urban Space Utilization with Renewable Wireless Sensor Network	133
<i>Billy Pik Lik Lau, Tanmay Chaturvedi, Benny Kai Kiat Ng, Kai Li, Marakkalage Sumudu Hasala, and Chau Yuen</i>	
Spatial Big Data for Designing Large Scale Infrastructure: A Case-Study of Electrical Road Systems	143
<i>Vinutha Magal Shreenath and Sebastiaan Meijer</i>	

Session 7: Scalability & Performance

An Improved Incremental Training Approach for Large Scaled Dataset Based on Support Vector Machine	149
<i>Jingcai Guo</i>	
SAPAM: A Scalable "Activities in Places" Analysis Mechanism for Informed Place Design	158
<i>Linlin You and Bige Tunçer</i>	
A Real-Time Big Data Analysis Framework on a CPU/GPU Heterogeneous Cluster: A Meteorological Application Case Study	168
<i>Mohamed Hassaan and Iman Elghandour</i>	

Session 8: Pattern Detection & Recognition

A Benchmarking Platform for Analyzing Corpora of Traces: The Recognition of the Users' Involvement in Fields of Competencies	178
<i>Christophe Courtin and Miguel Tomasena</i>	
A Big Data Analytics Based Approach to Anomaly Detection	187
<i>Abdul Razaq, Huaglory Tianfield, and Peter Barrie</i>	
Not Too Late to Identify Potential Churners: Early Churn Prediction in Telecommunication Industry	194
<i>Jingjiao Zhang, Jiaqing Fu, Chunhong Zhang, Xin Ke, and Zheng Hu</i>	

Session 9: Visual and Graph Analytics

Penalized Graph Partitioning Based Allocation Strategy for Database-as-a-Service Systems	200
<i>Tim Kiefer, Dirk Habich, and Wolfgang Lehner</i>	

Optimizing the Shortest Path Query on Large-Scale Dynamic Directed Graph	N/A
<i>Phuong-Hanh Du, Hai-Dang Pham, and Ngoc-Hoa Nguyen</i>	
Clustering Spatial Data by the Neighbors Intersection and the Density Difference	217
<i>Zhenglong Yan, Wenjian Luo, Chenyang Bu, and Li Ni</i>	

Session 10: Memory & Storage

NVHT: An Efficient Key-Value Storage Library for Non-Volatile Memory	227
<i>Jie Zhou, Yanyan Shen, Sumin Li, and Linpeng Huang</i>	
Node Architecture Implications for In-Memory Data Analytics on Scale-in Clusters	237
<i>Ahsan Javed Awan, Mats Brorsson, Vladimir Vlassov, and Eduard Ayguade</i>	
A Scalable Storage System for Structured Data Based on Higher Order Index Array	247
<i>Mehnuma Tabassum Omar and K.M. Azharul Hasan</i>	

Session 11: Big Data Applications 1

Neighborhood Features Help Detecting Non-Technical Losses in Big Data Sets	253
<i>Patrick Glauner, Jorge Augusto Meira, Lautaro Dolberg, Radu State, Franck Bettinger, and Yves Rangoni</i>	
Research on Semantic Orientation Classification of Chinese Online Product Reviews Based on Multi-Aspect Sentiment Analysis	262
<i>Qing Sun, Jianwei Niu, Zhong Yao, and Dongmin Qiu</i>	
Survey of Data Intensive Computing Technologies Application to Security Log Data Management	N/A
<i>Anne Tall, Jun Wang, and Dezhi Han</i>	
Social Community Detection Based on Node Distance and Interest	274
<i>Mohammed Ba Hutair, Ibrahim Kamel, and Zaher Al Agbari</i>	

Session 12: Big Data Applications 2

Non-Negative Multiple Matrix Factorization with Social Similarity for Recommender Systems	280
<i>Guoying Zhang, Min He, Hao Wu, Guanghui Cai, and Jianhong Ge</i>	
A Hybrid Method of Recommending POIs Based on Context and Personal Preference Confidence	287
<i>Jian Li, Guanjin Liu, Changjun Jiang, and Chungang Yan</i>	
Synergy and Antagonism in Online Advertising	293
<i>Jie Hou, Ya Zhang, and Xiao Gu</i>	

Behavior Profiling for Mobile Advertising	302
<i>Manxing Du, Radu State, Mats Brorsson, and Tigran Avenesov</i>	
Author Index	308