

2016 IEEE International Conference on Big Data (Big Data 2016)

**Washington, DC, USA
5-8 December 2016**

Pages 1-792



**IEEE Catalog Number: CFP16BGD-POD
ISBN: 978-1-4673-9006-4**

**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 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:	CFP16BGD-POD
ISBN (Print-On-Demand):	978-1-4673-9006-4
ISBN (Online):	978-1-4673-9005-7

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

TABLE OF CONTENTS

COGNITIVE COMPUTING: FROM BREAKTHROUGHS IN THE LAB TO APPLICATIONS ON THE FIELD	1
<i>Guruduth S. Banavar</i>	
HARNESSING THE DATA REVOLUTION: A PERSPECTIVE FROM THE NATIONAL SCIENCE FOUNDATION	2
<i>Chaitanya Baru</i>	
BIG DATA SECURITY AND PRIVACY	3
<i>Elisa Bertino</i>	
ON THE POWER OF BIG DATA: MINING STRUCTURES FROM MASSIVE, UNSTRUCTURED TEXT DATA	4
<i>Jiawei Han</i>	
LEVERAGING HIGH PERFORMANCE COMPUTING TO DRIVE ADVANCED MANUFACTURING R&D AT THE US DEPARTMENT OF ENERGY	5
<i>Mark Johnson</i>	
DATABASE DECAY AND HOW TO AVOID IT	7
<i>Michael Stonebraker ; Dong Deng ; Michael L. Brodie</i>	
CACHE-OBVIOUS LOOPS BASED ON A NOVEL SPACE-FILLING CURVE	17
<i>Christian Böhm ; Martin Perdacher ; Claudia Plant</i>	
DD-RTREE: A DYNAMIC DISTRIBUTED DATA STRUCTURE FOR EFFICIENT DATA DISTRIBUTION AMONG CLUSTER NODES FOR SPATIAL DATA MINING ALGORITHMS	27
<i>Jagat Sesh Challa ; Poonam Goyal ; S. Nikhil ; Aditya Mangla ; Sundar S. Balasubramaniam ; Navneet Goyal</i>	
A META-GRAPH APPROACH TO ANALYZE SUBGRAPH-CENTRIC DISTRIBUTED PROGRAMMING MODELS	37
<i>Ravikant Dindokar ; Neel Choudhury ; Yogesh Simmhan</i>	
EXACT STRUCTURE LEARNING OF BAYESIAN NETWORKS BY OPTIMAL PATH EXTENSION	48
<i>Subhadeep Karan ; Jaroslaw Zola</i>	
DATALOGRAPHY: SCALING DATALOG GRAPH ANALYTICS ON GRAPH PROCESSING SYSTEMS	56
<i>Walaa Eldin Moustafa ; Vicky Papavasileiou ; Ken Yocum ; Alin Deutsch</i>	
PREDICTING STATISTICS OF ASYNCHRONOUS SGD PARAMETERS FOR A LARGE-SCALE DISTRIBUTED DEEP LEARNING SYSTEM ON GPU SUPERCOMPUTERS	66
<i>Yosuke Oyama ; Akihiro Nomura ; Ikuro Sato ; Hiroki Nishimura ; Yukimasa Tamatsu ; Satoshi Matsuoka</i>	
CONSENSUS OPTIMIZATION WITH DELAYED AND STOCHASTIC GRADIENTS ON DECENTRALIZED NETWORKS	76
<i>Benjamin Sirb ; Xiaojing Ye</i>	
PAIRWISE TOPIC MODEL AND ITS APPLICATION TO TOPIC TRANSITION AND EVOLUTION	86
<i>Xiaoli Song ; Yan Rui ; Xiaohua Hu</i>	
INTERPRETABLE AND EFFECTIVE OPINION SPAM DETECTION VIA TEMPORAL PATTERNS MINING ACROSS WEBSITES	96
<i>Yuan Yuan ; Sihong Xie ; Chun-Ta Lu ; Jie Tang ; Philip S. Yu</i>	
A FAST STRUCTURED REGRESSION FOR LARGE NETWORKS	106
<i>Fang Zhou ; Mohamed Ghalwash ; Zoran Obradovic</i>	
ANTECEDENTS OF BIG DATA QUALITY: AN EMPIRICAL EXAMINATION IN FINANCIAL SERVICE ORGANIZATIONS	116
<i>Adiska Fardani Haryadi ; Joris Hulstijn ; Agung Wahyudi ; Haiko Van Der Voort ; Marijn Janssen</i>	
PSH: A PROBABILISTIC SIGNATURE HASH METHOD WITH HASH NEIGHBORHOOD CANDIDATE GENERATION FOR FAST EDIT-DISTANCE STRING COMPARISON ON BIG DATA	122
<i>Joseph Jupin ; Justin Y. Shi ; Eduard C. Dragut</i>	
EFFICIENT MULTIPLE SCALE KERNEL CLASSIFIERS	128
<i>Rocco Langone ; Johan A. K. Suykens</i>	
A THEORETICAL MODEL FOR N-GRAM DISTRIBUTION IN BIG DATA CORPORA	134
<i>Joaquim F. Silva ; Carlos Goncalves ; Jose C. Cunha</i>	

THE SELF-AVOIDING WALK-JUMP (SAWJ) ALGORITHM FOR FINDING MAXIMUM DEGREE NODES IN LARGE GRAPHS	142
<i>Jonathan Stokes ; Steven Weber</i>	
SEMANTIC PATTERN MINING FOR TEXT MINING	150
<i>Xiaoli Song ; Xiaotong Wang ; Xiaohua Hu</i>	
DETECTING GRADUAL CHANGES FROM DATA STREAM USING MDL-CHANGE STATISTICS	156
<i>Kenji Yamanishi ; Kohei Miyaguchi</i>	
EXPLOITING TEMPORAL DIVERGENCE OF TOPIC DISTRIBUTIONS FOR EVENT DETECTION	164
<i>Rongda Zhu ; Aston Zhang ; Jian Peng ; Chengxiang Zhai</i>	
THRILL: HIGH-PERFORMANCE ALGORITHMIC DISTRIBUTED BATCH DATA PROCESSING WITH C++	172
<i>Timo Bingmann ; Michael Axtmann ; Emanuel Jöbstl ; Sebastian Lamm ; Huyen Chau Nguyen ; Alexander Noe ; Sebastian Schlag ; Matthias Stumpp ; Tobias Sturm ; Peter Sanders</i>	
TOWARDS RESOURCE-EFFICIENT CLOUD SYSTEMS: AVOIDING OVER-PROVISIONING IN DEMAND-PREDICTION BASED RESOURCE PROVISIONING	184
<i>Lihua Chen ; Haiying Shen</i>	
MIX 'N' MATCH MULTI-ENGINE ANALYTICS	194
<i>Katerina Doka ; Nikolaos Papailiou ; Victor Giannakouris ; Dimitrios Tsoumakos ; Nectarios Koziris</i>	
MATRIX FACTORIZATIONS AT SCALE: A COMPARISON OF SCIENTIFIC DATA ANALYTICS IN SPARK AND C+MPI USING THREE CASE STUDIES	204
<i>Alex Gittens ; Aditya Devarakonda ; Evan Racadh ; Michael Ringenburt ; Lisa Gerhardt ; Jey Kottalam ; Jialin Liu ; Kristyn Maschhoff ; Shane Canon ; Jatin Chhugani ; Pramod Sharma ; Jiyan Yang ; James Demmel ; Jim Harrell ; Venkat Krishnamurthy ; Michael W. Mahoney ; Prabhat</i>	
YINMEM: A DISTRIBUTED PARALLEL INDEXED IN-MEMORY COMPUTATION SYSTEM FOR LARGE SCALE DATA ANALYTICS	214
<i>Yin Huang ; Yelena Yesha ; Milton Halem ; Yaacov Yesha ; Shujia Zhou</i>	
EFFICIENT DATA ACCESS STRATEGIES FOR HADOOP AND SPARK ON HPC CLUSTER WITH HETEROGENEOUS STORAGE	223
<i>Nusrat Sharmin Islam ; Md. Wasi-Ur-Rahman ; Xiaoyi Lu ; Dhabaleswar K. D K Panda</i>	
COMPARING APPLICATION PERFORMANCE ON HPC-BASED HADOOP PLATFORMS WITH LOCAL STORAGE AND DEDICATED STORAGE	233
<i>Zhuozhao Li ; Haiying Shen ; Jeffrey Denton ; Walter Ligon</i>	
CCRP: CUSTOMIZED COOPERATIVE RESOURCE PROVISIONING FOR HIGH RESOURCE UTILIZATION IN CLOUDS	243
<i>Jinwei Liu ; Haiying Shen ; Husnu S. Narman</i>	
HIGH-PERFORMANCE DESIGN OF APACHE SPARK WITH RDMA AND ITS BENEFITS ON VARIOUS WORKLOADS	253
<i>Xiaoyi Lu ; Dipti Shankar ; Shashank Gugnani ; Dhabaleswar K. D K Panda</i>	
A LOW-LOAD STREAM PROCESSING SCHEME FOR IOT ENVIRONMENTS	263
<i>Tomoki Yoshihisa ; Takahiro Hara</i>	
SPARK-GPU: AN ACCELERATED IN-MEMORY DATA PROCESSING ENGINE ON CLUSTERS	273
<i>Yuan Yuan ; Meisam Fathi Salmi ; Yin Huai ; Kaibo Wang ; Rubao Lee ; Xiaodong Zhang</i>	
ARGO: ARCHITECTURE-AWARE GRAPH PARTITIONING	284
<i>Angen Zheng ; Alexandros Labrinidis ; Panos K. Chrysanthis ; Jack Lange</i>	
ADAPTING TO DATA SPARSITY FOR EFFICIENT PARALLEL PARAFAC TENSOR DECOMPOSITION IN HADOOP	294
<i>Kareem S. Aggour ; Bilent Yener</i>	
CLOUD KOTTA: ENABLING SECURE AND SCALABLE DATA ANALYTICS IN THE CLOUD	302
<i>Yadu N. Babuji ; Kyle Chard ; Aaron Gerow ; Eamon Duede</i>	
ENTITY RESOLUTION ACCELERATION USING THE AUTOMATA PROCESSOR	311
<i>Chunkun Bo ; Ke Wang ; Jeffrey J. Fox ; Kevin Skadron</i>	
I'LL TAKE THAT TO GO: BIG DATA BAGS AND MINIMAL IDENTIFIERS FOR EXCHANGE OF LARGE, COMPLEX DATASETS	319
<i>Kyle Chard ; Mike D'Arcy ; Ben Heavner ; Ian Foster ; Carl Kesselman ; Ravi Madduri ; Alexis Rodriguez ; Stian Soiland-Reyes ; Carole Goble ; Kristi Clark ; Eric W. Deutsch ; Ivo Dinov ; Nathan Price ; Arthur Toga</i>	
MASSIVE PARALLELISM FOR NON-LINEAR AND NON-STATIONARY DATA ANALYSIS WITH GPGPU	329
<i>Chun-Chieh Chen ; Chih-Ya Shen ; Ming-Syan Chen</i>	
BIG DATA FRAMEWORK INTERFERENCE IN RESTRICTED PRIVATE CLOUD SETTINGS	335
<i>Stratos Dimopoulos ; Chandra Krintz ; Rich Wolski</i>	

EVALUATING THE IMPACT OF DATA PLACEMENT TO SPARK AND SCIDB WITH AN EARTH SCIENCE USE CASE	341
<i>Khoa Doan ; Amidu O Oloso ; Kwo-Sen Kuo ; Thomas L Clune ; Hongfeng Yu ; Brian Nelson ; Jian Zhang</i>	
JAVA THREAD AND PROCESS PERFORMANCE FOR PARALLEL MACHINE LEARNING ON MULTICORE HPC CLUSTERS	347
<i>Saliya Ekanayake ; Supun Kamburugamuve ; Pulasthi Wickramasinghe ; Geoffrey C. Fox</i>	
POWER EFFICIENT BIG DATA ANALYTICS ALGORITHMS THROUGH LOW-LEVEL OPERATIONS	355
<i>Gheorghii Guzun ; Josiah C. McClurg ; Guadalupe Cañahuate ; Raghuraman Mudumbai</i>	
EVALUATING THE IMPACTS OF CODE-LEVEL PERFORMANCE TUNINGS ON POWER EFFICIENCY	362
<i>Satoshi Imamura ; Keitaro Oka ; Yuichiro Yasui ; Yuichi Inadomi ; Katsuki Fujisawa ; Toshio Endo ; Koji Ueno ; Keiichiro Fukazawa ; Nozomi Hata ; Yuta Kakibuka ; Koji Inoue ; Takatsugu Ono</i>	
RADU: BRIDGING THE DIVIDE BETWEEN DATA AND INFRASTRUCTURE MANAGEMENT TO SUPPORT DATA-DRIVEN COLLABORATIONS	370
<i>Fan Jiang ; Claris Castillo ; Charles Schmitt</i>	
A COMPARISON OF GENERAL-PURPOSE DISTRIBUTED SYSTEMS FOR DATA PROCESSING	378
<i>Jinfeng Li ; James Cheng ; Yunjian Zhao ; Fan Yang ; Yuzhen Huang ; Haipeng Chen ; Ruihao Zhao</i>	
A POPULARITY-AWARE COST-EFFECTIVE REPLICATION SCHEME FOR HIGH DATA DURABILITY IN CLOUD STORAGE	384
<i>Jinwei Liu ; Haiying Shen</i>	
MANAGING HOT METADATA FOR SCIENTIFIC WORKFLOWS ON MULTISITE CLOUDS	390
<i>Luis Pineda-Morales ; Ji Liu ; Alexandru Costan ; Esther Pacitti ; Gabriel Antoniu ; Patrick Valduriez ; Marta Mattoso</i>	
I/O CHUNKING AND LATENCY HIDING APPROACH FOR OUT-OF-CORE SORTING ACCELERATION USING GPU AND FLASH NVM	398
<i>Hitoshi Sato ; Ryo Mizote ; Satoshi Matsuoka ; Hirotaka Ogawa</i>	
BOLDIO: A HYBRID AND RESILIENT BURST-BUFFER OVER LUSTRE FOR ACCELERATING BIG DATA I/O	404
<i>Dipti Shankar ; Xiaoyi Lu ; Dhabaleswar K. D K Panda</i>	
REAL TIME PROCESSING OF STREAMING AND STATIC INFORMATION	410
<i>Christoforos Svingos ; Theofilos Mailis ; Herald Kllapi ; Lefteris Stamatogiannakis ; Yannis Kotidis ; Yannis Ioannidis</i>	
HPTA: HIGH-PERFORMANCE TEXT ANALYTICS	416
<i>Hans Vandierendonck ; Karen Murphy ; Mahwish Arif ; Dimitrios S. Nikolopoulos</i>	
PERFORMANCE EVALUATION OF BIG DATA FRAMEWORKS FOR LARGE-SCALE DATA ANALYTICS	424
<i>Jorge Veiga ; Roberto R. Expósito ; Xoan C. Pardo ; Guillermo L. Taboada ; Juan Tourifio</i>	
SLA-BASED PROFIT OPTIMIZATION FOR RESOURCE MANAGEMENT OF BIG DATA ANALYTICS-AS-A-SERVICE PLATFORMS IN CLOUD COMPUTING ENVIRONMENTS	432
<i>Yali Zhao ; Rodrigo N. Calheiros ; James Bailey ; Richard Sinnott</i>	
MATERIALIZED VIEW SELECTION IN FEED FOLLOWING SYSTEMS	442
<i>Kaiji Chen ; Yongluan Zhou</i>	
MUSQLE: DISTRIBUTED SQL QUERY EXECUTION OVER MULTIPLE ENGINE ENVIRONMENTS	452
<i>Victor Giannakouris ; Nikolaos Papailiou ; Dimitrios Tsoumakos ; Nectarios Koziris</i>	
SAMPLING-BASED DISTRIBUTED KERNEL MEAN MATCHING USING SPARK	462
<i>Ahsanul Haque ; Zhuoyi Wang ; Swarup Chandra ; Yupeng Gao ; Latifur Khan ; Charu Aggarwal</i>	
CLOCKWISE COMPRESSION FOR TRAJECTORY DATA UNDER ROAD NETWORK CONSTRAINTS	472
<i>Yudian Ji ; Yuda Zang ; Wuman Luo ; Xibo Zhou ; Ye Ding ; Lionel M. Ni</i>	
SEMANTIC APPROACH TO AUTOMATING MANAGEMENT OF BIG DATA PRIVACY POLICIES	482
<i>Karuna P Joshi ; Aditi Gupta ; Sudip Mittal ; Claudia Pearce ; Anupam Joshi ; Tim Finin</i>	
HANDLING UNCERTAINTY IN TRAJECTORIES OF MOVING OBJECTS IN UNCONSTRAINED OUTDOOR SPACES	492
<i>Eleazar Leal ; Le Gruenwald ; Jianing Zhang</i>	
ACCELERATING RANGE QUERIES FOR LARGE-SCALE UNSTRUCTURED MESHES	502
<i>Cuong Nguyen ; Philip J. Rhodes</i>	
IN PURSUIT OF OUTLIERS IN MULTI-DIMENSIONAL DATA STREAMS	512
<i>Shiblee Sadik ; Le Gruenwald ; Eleazar Leal</i>	

WISDOM: WEIGHTED INCREMENTAL SPATIO-TEMPORAL MULTI-TASK LEARNING VIA TENSOR DECOMPOSITION	522
<i>Jianpeng Xu ; Jiayu Zhou ; Pang-Ning Tan ; Xi Liu ; Lifeng Luo</i>	
ADVANTAGE OF INTEGRATION IN BIG DATA: FEATURE GENERATION IN MULTI-RELATIONAL DATABASES FOR IMBALANCED LEARNING	532
<i>Farrukh Ahmed ; Michele Samorani ; Colin Bellinger ; Osmar R. Zaiane</i>	
SAMPLING LABELLED PROFILE DATA FOR IDENTITY RESOLUTION	540
<i>Matthew Edwards ; Stephen Wattam ; Paul Rayson ; Awais Rashid</i>	
PICK YOUR CHOICE IN HBASE: SECURITY OR PERFORMANCE	548
<i>Frank Pallas ; Johannes Günther ; David Bermbach</i>	
BDTUNE: HIERARCHICAL CORRELATION-BASED PERFORMANCE ANALYSIS AND RULE-BASED DIAGNOSIS FOR BIG DATA SYSTEMS	555
<i>Rui Ren ; Zhen Jia ; Lei Wang ; Jianfeng Zhan ; Tianxu Yi</i>	
TRANSFER LEARNING ALGORITHMS FOR AUTONOMOUS RECONFIGURATION OF WEARABLE SYSTEMS	563
<i>Ramyar Saeedi ; Hassan Ghasemzadeh ; Assefaw H. Gebremedhin</i>	
EFFICIENT PROCESSING OF TOP-K JOINS IN MAPREDUCE	570
<i>Mei Saouk ; Christos Doukeridis ; Akrivi Vlachou ; Kjetil Norvåg</i>	
OBJECT IDENTIFICATION WITH PAY-AS-YOU-GO CROWDSOURCING	578
<i>Ting Wu ; Chen Jason Zhang ; Lei Chen ; Pan Hui ; Siyuan Liu</i>	
ESTIMATION OF LOCAL SUBGRAPH COUNTS	586
<i>Nesreen K. Ahmed ; Theodore L. Willke ; Ryan A. Rossi</i>	
MULTI-STEP THRESHOLD ALGORITHM FOR EFFICIENT FEATURE-BASED QUERY PROCESSING IN LARGE-SCALE MULTIMEDIA DATABASES	596
<i>Christian Beecks ; Alexander Graß</i>	
PRIIME: A GENERIC FRAMEWORK FOR INTERACTIVE PERSONALIZED INTERESTING PATTERN DISCOVERY	606
<i>Mansurul A Bhuiyan ; Mohammad Al Hasan</i>	
LABELING ACTORS IN MULTI-VIEW SOCIAL NETWORKS BY INTEGRATING INFORMATION FROM WITHIN AND ACROSS MULTIPLE VIEWS	616
<i>Ngot Bui ; Thanh Le ; Vasant Honavar</i>	
ONLINE SOCIAL NETWORK EVOLUTION: REVISITING THE TWITTER GRAPH	626
<i>Hariton Efstathiades ; Demetris Antoniadis ; George Pallis ; Marios D. Dikaiakos ; Zoltán Szlávik ; Robert-Jan Sips</i>	
PARALLEL TOP-K SUBGRAPH QUERY IN MASSIVE GRAPHS: COMPUTING FROM THE PERSPECTIVE OF SINGLE VERTEX	636
<i>Jianliang Gao ; Bo Song ; Ping Liu ; Weimao Ke ; Jianxin Wang ; Xiaohua Hu</i>	
REQUEST: A SCALABLE FRAMEWORK FOR INTERACTIVE CONSTRUCTION OF EXPLORATORY QUERIES	646
<i>Xiaoyu Ge ; Yanbing Xue ; Zhipeng Luo ; Mohamed A. Sharaf ; Panos K. Chrysanthis</i>	
DYNAMIC FEATURE GENERATION AND SELECTION ON HETEROGENEOUS GRAPH FOR MUSIC RECOMMENDATION	656
<i>Chun Guo ; Xiaozhong Liu</i>	
AN ADAPTIVE INFORMATION-THEORETIC APPROACH FOR IDENTIFYING TEMPORAL CORRELATIONS IN BIG DATA SETS	666
<i>Nguyen Ho ; Huy Vo ; Mai Vu</i>	
TOWARDS UNSUPERVISED HOME LOCATION INFERENCE FROM ONLINE SOCIAL MEDIA	676
<i>Chao Huang ; Dong Wang ; Shenglong Zhu ; Daniel Yue Zhang</i>	
IMPROVED METHODS FOR STATIC INDEX PRUNING	686
<i>Wei Jiang ; Juan Rodriguez ; Torsten Suel</i>	
PARALLEL COMPUTATION OF K-NEAREST NEIGHBOR JOINS USING MAPREDUCE	696
<i>Wooyeol Kim ; Younghoon Kim ; Kyuseok Shim</i>	
HARNESSING RELATIONSHIPS FOR DOMAIN-SPECIFIC SUBGRAPH EXTRACTION: A RECOMMENDATION USE CASE	706
<i>Sarasi Lalithsena ; Pavan Kapanipathi ; Amit Sheth</i>	
SCALABLE LINK COMMUNITY DETECTION: A LOCAL DISPERSION-AWARE APPROACH	716
<i>Panagiotis Liakos ; Alexandros Ntoulas ; Alex Delis</i>	
OUTLIER DETECTION VIA SAMPLING ENSEMBLE	726
<i>Hongfu Liu ; Yuchao Zhang ; Bo Deng ; Yun Fu</i>	
RANDOM SURFING ON MULTIPARTITE GRAPHS	736
<i>Athanasios N. Nikolakopoulos ; Antonia Korba ; John D. Garofalakis</i>	

AN ACTIVE LEARNING METHOD FOR DATA STREAMS WITH CONCEPT DRIFT	746
<i>Cheong Hee Park ; Youngsoon Kang</i>	
ADAPTIVE NEURON APOPTOSIS FOR ACCELERATING DEEP LEARNING ON LARGE SCALE SYSTEMS	753
<i>Charles Siegel ; Jeff Daily ; Abhinav Vishnu</i>	
DELTASHERLOCK: IDENTIFYING CHANGES IN THE CLOUD	763
<i>Ata Turk ; Hao Chen ; Anthony Byrne ; John Knollmeyer ; Sastry S. Duri ; Canturk Isci ; Ayse K. Coskun</i>	
COMMUNITY DETECTION WITH PARTIALLY OBSERVABLE LINKS AND NODE ATTRIBUTES	773
<i>Xiaokai Wei ; Bokai Cao ; Weixiang Shao ; Chun-Ta Lu ; Philip S. Yu</i>	
PARALLEL GATHERING DISCOVERY OVER BIG TRAJECTORY DATA	783
<i>Yongyi Xian ; Yan Liu ; Chuanfei Xu</i>	
CER: COMPLEMENTARY ENTITY RECOGNITION VIA KNOWLEDGE EXPANSION ON LARGE UNLABELED PRODUCT REVIEWS	793
<i>Hu Xu ; Sihong Xie ; Lei Shu ; Philip S. Yu</i>	
HEER: HETEROGENEOUS GRAPH EMBEDDING FOR EMERGING RELATION DETECTION FROM NEWS	803
<i>Jingyuan Zhang ; Chun-Ta Lu ; Mianwei Zhou ; Sihong Xie ; Yi Chang ; Philip S. Yu</i>	
EFFICIENT TRIANGLE LISTING FOR BILLION-SCALE GRAPHS	813
<i>Hao Zhang ; Yuanyuan Zhu ; Lu Qin ; Hong Cheng ; Jeffrey Xu Yu</i>	
TOWARDS UNDERSTANDING WORD EMBEDDINGS: AUTOMATICALLY EXPLAINING SIMILARITY OF TERMS	823
<i>Yating Zhang ; Adam Jatowt ; Katsumi Tanaka</i>	
PREDICTING TAXI DEMAND AT HIGH SPATIAL RESOLUTION: APPROACHING THE LIMIT OF PREDICTABILITY	833
<i>Kai Zhao ; Denis Khryashchev ; Juliana Freire ; Cláudio Silva ; Huy Vo</i>	
TELCOFLOW: VISUAL EXPLORATION OF COLLECTIVE BEHAVIORS BASED ON TELCO DATA	843
<i>Yixian Zheng ; Wenchao Wu ; Haipeng Zeng ; Nan Cao ; Huamin Qu ; Mingxuan Yuan ; Jia Zeng ; Lionel M. Ni</i>	
DISTRIBUTED AND PARALLEL HIGH UTILITY SEQUENTIAL PATTERN MINING	853
<i>Morteza Zihayat ; Zane Zhenhua Hut ; Aijun An ; Yonggang Hut</i>	
IMPROVING EFFICIENCY OF MAXIMIZING SPREAD IN THE FLOW AUTHORITY MODEL FOR LARGE SPARSE NETWORKS	863
<i>Philip K. Chan ; Ebad Ahmadzadeh</i>	
SEMI-SUPERVISED DIRICHLET-HAWKES PROCESS WITH APPLICATIONS OF TOPIC DETECTION AND TRACKING IN TWITTER	869
<i>Wanying Ding ; Yue Zhang ; Chaomei Chen ; Xiaohua Hu</i>	
EFFECTIVE AND EFFICIENT GRAPH AUGMENTATION IN LARGE GRAPHS	875
<i>Ioanna Filippidou ; Yannis Kotidis</i>	
FAST NEAREST NEIGHBOR SEARCH THROUGH SPARSE RANDOM PROJECTIONS AND VOTING	881
<i>Ville Hyvönen ; Teemu Pitkänen ; Sotiris Tasoulis ; Elias Jääsaari ; Risto Tuomainen ; Liang Wang ; Jukka Corander ; Teemu Roos</i>	
SUMMARIZING BIG GRAPHS BY MEANS OF PSEUDO-BOOLEAN CONSTRAINTS	889
<i>Said Jabbour ; Nizar Mhadhbi ; Abdesattar Mhadhbi ; Badran Radaoui ; Lakhdar Sais</i>	
BIG DATA ON A FEW PIXELS	895
<i>Uwe Jugel ; Zbigniew Jerzak ; Volker Markl</i>	
SHAPE MATCHING USING SKELETON CONTEXT FOR AUTOMATED BOW ECHO DETECTION	901
<i>Mohammad Mahdi Kamani ; Farshid Farhat ; Stephen Wistar ; James Z. Wang</i>	
SCALABILITY ANALYSIS OF DISTRIBUTED SEARCH IN LARGE PEER-TO-PEER NETWORKS	909
<i>Weimao Ke ; Javed Mostafa</i>	
VHT: VERTICAL Hoeffding Tree	915
<i>Nicolas Kourtellis ; Gianmarco De Francisci Morales ; Albert Bifet ; Arinto Murdopo</i>	
COMPRESSED LEARNING FOR TIME SERIES CLASSIFICATION	923
<i>Yuh-Jye Lee ; Hsing-Kuo Pao ; Shueh-Han Shih ; Jing-Yao Lin ; Xin-Rong Chen</i>	
CONNECTION DISCOVERY USING SHARED IMAGES BY GAUSSIAN RELATIONAL TOPIC MODEL	931
<i>Xiaopeng Li ; Ming Cheung ; James She</i>	

INFERRING RESTAURANT STYLES BY MINING CROWD SOURCED PHOTOS FROM USER-REVIEW WEBSITES	937
<i>Haofu Liao ; Yucheng Li ; Tianran Hu ; Jiebo Luo</i>	
MULTIPLE SUBMODELS PARALLEL SUPPORT VECTOR MACHINE ON SPARK	945
<i>Chang Liu ; Bin Wu ; Yi Yang ; Zhihong Guo</i>	
WHAT MAKES A GROUP FAIL: MODELING SOCIAL GROUP BEHAVIOR IN EVENT-BASED SOCIAL NETWORKS	951
<i>Xiang Liu ; Torsten Suel</i>	
EFFICIENT LARGE SCALE NEAR-DUPLICATE VIDEO DETECTION BASE ON SPARK	957
<i>Jinna Lv ; Bin Wu ; Shuai Yang ; Bingjing Jia ; Peigang Qiu</i>	
CONTEXT-AWARE POINT OF INTEREST RECOMMENDATION USING TENSOR FACTORIZATION	963
<i>Stathis Maroulis ; Ioannis Boutsis ; Vana Kalogeraki</i>	
PERSISTENT CASCADES: MEASURING FUNDAMENTAL COMMUNICATION STRUCTURE IN SOCIAL NETWORKS	969
<i>Steven Morse ; Marta C. Gonzalez ; Natasha Markuzon</i>	
TRUTHCORE: NON-PARAMETRIC ESTIMATION OF TRUTH FROM A COLLECTION OF AUTHORITATIVE SOURCES	976
<i>Tathagata Mukherjee ; Biswas Parajuli ; Piyush Kumar ; Eduardo Pasillao</i>	
EFFICIENT INDEX UPDATES FOR MIXED UPDATE AND QUERY LOADS	984
<i>Sergey Nepomnyachiy ; Torsten Suel</i>	
COMPARTMENTALIZED ADAPTIVE TOPIC MINING ON SOCIAL MEDIA STREAMS	992
<i>Gopi Chand Nutakki ; Olfa Nasraoui</i>	
COMPUTING TRIANGLE AND OPEN-WEDGE HEAVY-HITTERS IN LARGE NETWORKS	998
<i>A. Pavan ; P. Quint ; S. Scott ; N. V. Vinodchandran ; J. Smith</i>	
ADDRESSING THE BIG-EARTH-DATA VARIETY CHALLENGE WITH THE HIERARCHICAL TRIANGULAR MESH	1006
<i>Michael L. Rilee ; Kwo-Sen Kuo ; Thomas Clune ; Amidu Olosoto ; Paul G. Brown ; Hongfeng Yu</i>	
ONLINE MULTI-VIEW CLUSTERING WITH INCOMPLETE VIEWS	1012
<i>Weixiang Shao ; Lifang He ; Chun-Ta Lu ; Philip S. Yu</i>	
EXPENDITURE AWARE RATING PREDICTION FOR RECOMMENDATION	1018
<i>Chuan Shi ; Bowei He ; Menghao Zhang ; Fuzhen Zhuang ; Philip S. Yu ; Naiwang Guo</i>	
KERNELS FOR SCALABLE DATA ANALYSIS IN SCIENCE: TOWARDS AN ARCHITECTURE-PORTABLE FUTURE	1026
<i>Sreenivas R. Sukumar ; Ramakrishnan Kannan ; Seung-Hwan Lim ; Michael A. Matheson</i>	
SCALABLE DYNAMIC GRAPH SUMMARIZATION	1032
<i>Ioanna Tsalouchidou ; Gianmarco De Francisci Morales ; Francesco Bonchi ; Ricardo Baeza-Yates</i>	
EXTREME SCALE BREADTH-FIRST SEARCH ON SUPERCOMPUTERS	1040
<i>Koji Ueno ; Toyotaro Suzumura ; Naoya Maruyama ; Katsuki Fujisawa ; Satoshi Matsuoka</i>	
THREE-HOP DISTANCE ESTIMATION IN SOCIAL GRAPHS	1048
<i>Pascal Welke ; Alexander Markowetz ; Torsten Suel ; Maria Christoforaki</i>	
INCREMENTAL LEARNING FOR MATRIX FACTORIZATION IN RECOMMENDER SYSTEMS	1056
<i>Tong Yu ; Ole J. Mengshoel ; Alvin Jude ; Eugen Feller ; Julien Forgeat ; Nimish Radia</i>	
PARALLEL CLUSTERING METHOD FOR NON-DISJOINT PARTITIONING OF LARGE-SCALE DATA BASED ON SPARK FRAMEWORK	1064
<i>Abir Zayani ; Chiheb-Eddine Ben N'Cir ; Nadia Essoussi</i>	
POINT OF INTEREST RECOMMENDATION WITH SOCIAL AND GEOGRAPHICAL INFLUENCE	1070
<i>Da-Chuan Zhang ; Mei Li ; Chang-Dong Wang</i>	
ON ROBUST TRUTH DISCOVERY IN SPARSE SOCIAL MEDIA SENSING	1076
<i>Daniel Yue Zhang ; Rungang Han ; Dong Wang ; Chao Huang</i>	
ON THE FEASIBILITY OF AN EMBEDDED MACHINE LEARNING PROCESSOR FOR INTRUSION DETECTION	1082
<i>Rajesh Sankaran ; Ricado A. Calix</i>	
ANDROID MALWARE DEVELOPMENT ON PUBLIC MALWARE SCANNING PLATFORMS: A LARGE-SCALE DATA-DRIVEN STUDY	1090
<i>Heqing Huang ; Cong Zheng ; Junyuan Zeng ; Wu Zhou ; Sencun Zhu ; Peng Liu ; Suresh Chari ; Ce Zhang</i>	
IMPROVING THE UTILITY IN DIFFERENTIAL PRIVATE HISTOGRAM PUBLISHING: THEORETICAL STUDY AND PRACTICE	1100
<i>Hui Li ; Jiangtao Cui ; Xiaobin Lin ; Jianfeng Ma</i>	

DISTSD: DISTANCE-BASED SOCIAL DISCOVERY WITH PERSONALIZED POSTERIOR SCREENING	1110
<i>Xiao Pan ; Jiawei Zhang ; Fengjiao Wang ; Philip S. Yu</i>	
H2O: A HYBRID AND HIERARCHICAL OUTLIER DETECTION METHOD FOR LARGE SCALE DATA PROTECTION	1120
<i>Quan Zhang ; Mu Qiao ; Ramani R. Routray ; Weisong Shi</i>	
SCALABLE ATTACK PROPAGATION MODEL AND ALGORITHMS FOR HONEYPOT SYSTEMS	1130
<i>Ariel Bar ; Bracha Shapira ; Lior Rokach ; Moshe Unger</i>	
LOCAL SUBSPACE-BASED OUTLIER DETECTION USING GLOBAL NEIGHBOURHOODS	1136
<i>Bas Van Stein ; Matthijs Van Leeuwen ; Thomas Bäck</i>	
PROTECTING THE LOCATION PRIVACY OF MOBILE SOCIAL MEDIA USERS	1143
<i>Shuo Wang ; Richard Sinnott ; Surya Nepal</i>	
ENABLING FACTOR ANALYSIS ON THOUSAND-SUBJECT NEUROIMAGING DATASETS	1151
<i>Michael J. Anderson ; Mihai Capota ; Javier S. Turek ; Xia Zhu ; Theodore L. Willke ; Yida Wang ; Po-Hsuan Chen ; Jeremy R. Manning ; Peter J. Ramadge ; Kenneth A. Norman</i>	
SHOOTING A MOVING TARGET: MOTION-PREDICTION-BASED TRANSMISSION FOR 360-DEGREE VIDEOS	1161
<i>Yanan Bao ; Huasen Wu ; Tianxiao Zhang ; Albara Ah Ramli ; Xin Liu</i>	
LAZER: DISTRIBUTED MEMORY-EFFICIENT ASSEMBLY OF LARGE-SCALE GENOMES	1171
<i>Sayan Goswami ; Arghya Kusum Das ; Richard Platania ; Kisung Lee ; Seung-Jong Park</i>	
LEVERAGING MULTI-GRANULARITY ENERGY DATA FOR ACCURATE ENERGY DEMAND FORECAST IN SMART GRIDS	1182
<i>Zhichuan Huang ; Ting Zhu</i>	
LEARNING LARGE-SCALE PLANTATION MAPPING FROM IMPERFECT ANNOTATORS	1192
<i>Xiaowei Jia ; Ankush Khandelwal ; James Gerber ; Kimberly Carlson ; Paul West ; Vipin Kumar</i>	
AD ALLOCATION WITH SECONDARY METRICS	1202
<i>Darja Krushevskaja ; William Simpson ; S. Muthukrishnan</i>	
EMBEDDING FEATURE SELECTION FOR LARGE-SCALE HIERARCHICAL CLASSIFICATION	1212
<i>Azad Naik ; Huzefa Rangwala</i>	
NETWORK ANALYSIS FOR IDENTIFYING AND CHARACTERIZING DISEASE OUTBREAK INFLUENCE FROM VOLUMINOUS EPIDEMIOLOGY DATA	1222
<i>Naman Shah ; Harshil Shah ; Matthew Malensek ; Sangmi Lee Pallickara ; Shrideep Pallickara</i>	
SCALABLE GENOMICS: FROM RAW DATA TO ALIGNED READS ON APACHE YARN	1232
<i>Francesco Versaci ; Luca Pireddu ; Gianluigi Zanetti</i>	
REAL-TIME FULL CORRELATION MATRIX ANALYSIS OF FMRI DATA	1242
<i>Yida Wang ; Bryn Keller ; Mihai Capota ; Michael J. Anderson ; Narayanan Sundaram ; Jonathan D. Cohen ; Kai Li ; Nicholas B. Turk-Browne ; Theodore L. Willke</i>	
WHEN REMOTE SENSING DATA MEET UBIQUITOUS URBAN DATA: FINE-GRAINED AIR QUALITY INFERENCE	1252
<i>Yanan Xu ; Yanmin Zhu</i>	
BUYER TARGETING OPTIMIZATION: A UNIFIED CUSTOMER SEGMENTATION PERSPECTIVE	1262
<i>Jingyuan Yang ; Chuanren Liu ; Mingfei Teng ; March Liao ; Hui Xiong</i>	
USING MACHINE LEARNING TO IDENTIFY MAJOR SHIFTS IN HUMAN GUT MICROBIOME PROTEIN FAMILY ABUNDANCE IN DISEASE	1272
<i>Mehrdad Yazdani ; Bryn C. Taylor ; Justine W. Debelius ; Weizhong Li ; Rob Knight ; Larry Smarr</i>	
ONLINE INFERENCE FOR TIME-VARYING TEMPORAL DEPENDENCY DISCOVERY FROM TIME SERIES	1281
<i>Chunqiu Zeng ; Qing Wang ; Wentao Wang ; Tao Li ; Larisa Shwartz</i>	
AUTOMATED IT SYSTEM FAILURE PREDICTION: A DEEP LEARNING APPROACH	1291
<i>Ke Zhang ; Jianwu Xu ; Martin Renqiang Min ; Guofei Jiang ; Konstantinos Pelechrinis ; Hui Zhang</i>	
ESTIMATING HUMAN INTERACTIONS WITH ELECTRICAL APPLIANCES FOR ACTIVITY-BASED ENERGY SAVINGS RECOMMENDATIONS	1301
<i>Hông-ân Cao ; Tri Kurniawan Wijaya ; Karl Aberer ; Nuno Nunes</i>	
SCALABLE NEAREST NEIGHBOR BASED HIERARCHICAL CHANGE DETECTION FRAMEWORK FOR CROP MONITORING	1309
<i>Zexi Chen ; Ranga Raju Vatsavai ; Bharathkumar Ramachandra ; Qiang Zhang ; Nagendra Singh ; Sreenivas Sukumar</i>	
OPTIMIZING CALLOUT IN UNIFIED AD MARKETS	1315
<i>Aman Gupta ; S. Muthukrishnan ; Smita Wadhwa</i>	

APPLICATION-DRIVEN SENSING DATA RECONSTRUCTION AND SELECTION BASED ON CORRELATION MINING AND DYNAMIC FEEDBACK	1322
<i>Zhichuan Huang ; Tiantian Xie ; Ting Zhu ; Jianwu Wang ; Qingquan Zhang</i>	
IDENTIFYING DYNAMIC CHANGES WITH NOISY LABELS IN SPATIAL-TEMPORAL DATA: A STUDY ON LARGE-SCALE WATER MONITORING APPLICATION.....	1328
<i>Xiaowei Jia ; Xi Chen ; Anuj Karpatne ; Vipin Kumar</i>	
A STRATEGIC APPROACH FOR VISUALIZING THE VALUE OF BIG DATA (SAVV-BIGD) FRAMEWORK	1334
<i>Mike Lakoju ; Alan Serrano</i>	
A SCALABLE APPROACH FOR LOCATION-SPECIFIC DETECTION OF SANTA ANA CONDITIONS	1340
<i>Mai H. Nguyen ; Dylan Uys ; Daniel Crawl ; Charles Cowart ; Ilkay Altintas</i>	
EXPERIENCES WITH SMART CITY TRAFFIC PILOT	1346
<i>Susanna Pirttikangas ; Ekaterina Gilman ; Xiang Su ; Teemu Leppänen ; Anja Keskinarkaus ; Mika Rautiainen ; Mikko Pyykkönen ; Jukka Riekkii</i>	
HOW INTERESTING IMAGES ARE: AN ATYPICALITY APPROACH FOR SOCIAL NETWORKS.....	1353
<i>Elyas Sabeti ; Anders Host-Madsen</i>	
EXPLORING MEMORY HIERARCHY AND NETWORK TOPOLOGY FOR RUNTIME AMR DATA SHARING ACROSS SCIENTIFIC APPLICATIONS	1359
<i>Wenzhao Zhang ; Houjun Tang ; Stephen Ranshous ; Surendra Byna ; Daniel F. Martín ; Kesheng Wu ; Bin Dong ; Scott Klasky ; Nagiza F. Samatova</i>	
PITFALLS OF LONG-TERM ONLINE CONTROLLED EXPERIMENTS.....	1367
<i>Pavel Dmitriev ; Brian Frasca ; Somit Gupta ; Ron Kohavi ; Garnet Vaz</i>	
AN ARCHITECTURE FOR THE DEPLOYMENT OF STATISTICAL MODELS FOR THE BIG DATA ERA	1377
<i>Juergen Heit ; Jiayi Liu ; Mohak Shah</i>	
INFORMATION RETRIEVAL, FUSION, COMPLETION, AND CLUSTERING FOR EMPLOYEE EXPERTISE ESTIMATION.....	1385
<i>Raya Horesh ; Kush R. Varshney ; Jinfeng Yi</i>	
EMPIRICAL EVALUATIONS OF PREPROCESSING PARAMETERS' IMPACT ON PREDICTIVE CODING'S EFFECTIVENESS	1394
<i>Rishi Chhatwal ; Nathaniel Huber-Fliflet ; Robert Keeling ; Jianping Zhang ; Haozhen Zhao</i>	
LOGPROV: LOGGING EVENTS AS PROVENANCE OF BIG DATA ANALYTICS PIPELINES WITH TRUSTWORTHINESS.....	1402
<i>Ruoyu Wang ; Daniel Sun ; Guoqiang Li ; Muhammad Atif ; Surya Nepal</i>	
PATTERN RECOGNITION AND CLASSIFICATION OF HVAC RULE-BASED FAULTS IN COMMERCIAL BUILDINGS	1412
<i>Bradford Littooy ; Sophie Loire ; Michael Georgescu ; Igor Mezic</i>	
DEEP PARALLELIZATION OF PARALLEL FP-GROWTH USING PARENT-CHILD MAPREDUCE	1422
<i>Adetokunbo Makanju ; Zahra Farzanyar ; Aijun An ; Nick Cercone ; Zane Zhenhua Hu ; Yonggang Hu</i>	
THE STATE OF SQL-ON-HADOOP IN THE CLOUD	1432
<i>Nicolas Poggi ; Josep Ll. Berral ; Thomas Fenech ; David Carrera ; José Blakeley ; Umar Farooq Minhas ; Nikola Vujic</i>	
DETECTING FRAUD, CORRUPTION, AND COLLUSION IN INTERNATIONAL DEVELOPMENT CONTRACTS: THE DESIGN OF A PROOF-OF-CONCEPT AUTOMATED SYSTEM.....	1444
<i>Emily Grace ; Ankit Rai ; Elissa Redmiles ; Rayid Ghani</i>	
AUTOMATIC GENERATION OF RELATIONAL ATTRIBUTES: AN APPLICATION TO PRODUCT RETURNS	1454
<i>Michele Samorani ; Farrukh Ahmed ; Osmar R. Zaiane</i>	
DATA-AT-REST SECURITY FOR SPARK	1464
<i>Syed Yousaf Shah ; Brent Paulovicks ; Petros Zerfos</i>	
DO WE TRUST IMAGE MEASUREMENTS? VARIABILITY, ACCURACY AND TRACEABILITY OF IMAGE FEATURES	1474
<i>Mylene Simon ; Joe Chalfoun ; Mary Brady ; Peter Bajcsy</i>	
MINI-APPS FOR HIGH PERFORMANCE DATA ANALYSIS.....	1483
<i>Sreenivas R. Sukumar ; Michael A. Matheson ; Ramakrishnan Kannan ; Seung-Hwan Lim</i>	
PREDICTING ANNUAL AVERAGE DAILY HIGHWAY TRAFFIC FROM LARGE DATA AND VERY FEW MEASUREMENTS.....	1493
<i>Tomasz Tajmajer ; Malwina Splawinska ; Piotr Wasilewski ; Stan Matwin</i>	

FAST, LENIENT AND ACCURATE: BUILDING PERSONALIZED INSTANT SEARCH EXPERIENCE AT LINKEDIN.....	1502
<i>Ganesh Venkataraman ; Abhimanyu Lad ; Lin Guo ; Shakti Sinha</i>	
DIVERSIFYING TRENDING TOPIC DISCOVERY VIA SEMIDEFINITE PROGRAMMING.....	1512
<i>Hui Wu ; Yi Fang ; Huming Wu ; Shenhong Zhu</i>	
STORYTELLING IN HETEROGENEOUS TWITTER ENTITY NETWORK BASED ON HIERARCHICAL CLUSTER ROUTING.....	1522
<i>Xuchao Zhang ; Zhiqian Chen ; Weisheng Zhong ; Arnold P. Boedihardjo ; Chang-Tien Lu</i>	
QUANTIFYING SKILL RELEVANCE TO JOB TITLES.....	1532
<i>Wenjun Zhou ; Yun Zhu ; Faizan Javed ; Mahmudur Rahman ; Janani Balaji ; Matt McNair</i>	
SMARTCACHE: APPLICATION LAYER CACHING TO IMPROVE PERFORMANCE OF LARGE-SCALE MEMORY MAPPING.....	1542
<i>Zhenyun Zhuang ; Haricharan Ramachandra ; Badri Sridharan ; Brandon Duncan ; Kishore Gopalakrishna ; Jean-Francois Im</i>	
HIDDEN MARKOV BASED ANOMALY DETECTION FOR WATER SUPPLY SYSTEMS.....	1551
<i>Zahra Zohrevand ; Uwe Glasser ; Hamed Yaghoubi Shahir ; Mohammad A. Tayebi ; Robert Costanzo</i>	
ADVANCING NLP VIA A DISTRIBUTED-MESSAGING APPROACH.....	1561
<i>Ilaria Bordino ; Andrea Ferretti ; Marco Firrincieli ; Francesco Gullo ; Marcello Paris ; Stefano Pascolutti ; Gianluca Sabena</i>	
AUTOMATED PORT TRAFFIC STATISTICS: FROM RAW DATA TO VISUALISATION.....	1569
<i>Luca Cazzanti ; Antonio Davoli ; Leonardo M. Millefiori</i>	
USTORE: AN OPTIMIZED STORAGE SYSTEM FOR ENTERPRISE DATA WAREHOUSES AT UNIONPAY.....	1574
<i>Hongfeng Chai ; Hao Liu ; Xibo Zhou ; Yanjun Xu ; Shuo He ; Jinzhi Hua ; Dongjie He ; Weihuai Liu</i>	
EXTENSIVE LARGE-SCALE STUDY OF ERROR SURFACES IN SAMPLING-BASED DISTINCT VALUE ESTIMATORS FOR DATABASES.....	1579
<i>Vinay Deolalikar ; Hernan Laffitte</i>	
FORECASTING SQUATTING OF DEMAND IN DISPLAY ADVERTISING.....	1587
<i>Amita Gajewar ; Lizhong Wu ; Jignesh Parmar ; Ramana Yerneni</i>	
DATA QUALITY: EXPERIENCES AND LESSONS FROM OPERATIONALIZING BIG DATA.....	1595
<i>Archana Ganapathi ; Yanpei Chen</i>	
KDD MEETS BIG DATA.....	1603
<i>Nancy W. Grady</i>	
CLASSIFICATION OF MASSIVE MOBILE WEB LOG URLS FOR CUSTOMER PROFILING & ANALYTICS.....	1609
<i>Rajaraman Kanagasabai ; Anitha Veeramani ; Hu Shangfeng ; Kajanan Sangaralingam ; Giuseppe Manai</i>	
COMPANY RECOMMENDATION FOR NEW GRADUATES VIA IMPLICIT FEEDBACK MULTIPLE MATRIX FACTORIZATION WITH BAYESIAN OPTIMIZATION.....	1615
<i>Masahiro Kazama ; Issei Sato ; Haruaki Yatabe ; Tairiku Ogihara ; Tetsuro Onishi ; Hiroshi Nakagawa</i>	
HUMAN NETWORK USAGE PATTERNS REVEALED BY TELECOM DATA.....	1621
<i>Yiming Kong ; Hui Zang ; Xiaoli Ma</i>	
A DISTRIBUTED APPROACH TO ESTIMATING SEA PORT OPERATIONAL REGIONS FROM LOTS OF AIS DATA.....	1627
<i>Leonardo M. Millefiori ; Dimitrios Zisis ; Luca Cazzanti ; Gianfranco Arcieri</i>	
UNIFORMIZATION, ORGANIZATION, ASSOCIATION AND USE OF METADATA FROM MULTIPLE CONTENT PROVIDERS AND MANUFACTURERS: A CLOSE LOOK AT THE BUILDING AUTOMATION SYSTEM (BAS) SECTOR.....	1633
<i>Thibaud Nesztler ; Don Kasper ; Michael Georgescu ; Sophie Loire ; Igor Mezic</i>	
QED: Groupon's ETL Management and Curated Feature Catalog System for Machine Learning.....	1639
<i>Derrick C. Spell ; Ling-Yong Wang ; Richard T. Shomer ; Bahador Nooraei ; Jarrell Waggoner ; Xiao-Han T. Zeng ; Jae Young Chung ; Kai-Chen Cheng ; Daniel Kirsche</i>	
BIG-DATA-DRIVEN ANOMALY DETECTION IN INDUSTRY (4.0): AN APPROACH AND A CASE STUDY.....	1647
<i>Ljiljana Stojanovic ; Marko Dinic ; Nenad Stojanovic ; Aleksandar Stojadinovic</i>	
CROSS-MODAL EVENT SUMMARIZATION: A NETWORK OF NETWORKS APPROACH.....	1653
<i>Jiejun Xu ; Samuel D. Johnson ; Kang-Yu Ni</i>	
MANAGING A COMPLICATED WORKFLOW BASED ON DATAFLOW-BASED WORKFLOW SCHEDULER.....	1658
<i>Teruyoshi Zenmyo ; Satoshi Iijima ; Ichiro Fukuda</i>	
AN EDGE-SET BASED LARGE SCALE GRAPH PROCESSING SYSTEM.....	1664
<i>Li Zhou ; Yinglong Xia ; Hui Zang ; Jian Xu ; Mingzhen Xia</i>	

EVENT DETECTION FROM SOCIAL NETWORK STREAMS USING FREQUENT PATTERN MINING WITH DYNAMIC SUPPORT VALUES	1670
<i>Nora Alkhamees ; Maria Fasli</i>	
BIG DATA ANALYTICS IN CLOUD GAMING: PLAYERS' PATTERNS RECOGNITION USING ARTIFICIAL NEURAL NETWORKS	1680
<i>Victor Perazzolo Barros ; Pollyana Notargiacomo</i>	
MAPREDUCE-BASED DEEP LEARNING WITH HANDWRITTEN DIGIT RECOGNITION CASE STUDY	1690
<i>Nada Basit ; Yutong Zhang ; Hao Wu ; Haoran Liu ; Jieming Bin ; Yijun He ; Abdeltawab M. Hendawi</i>	
TEXT MINING AND SENTIMENT EXTRACTION IN CENTRAL BANK DOCUMENTS	1700
<i>Giuseppe Bruno</i>	
TO LINK OR NOT TO LINK: RANKING HYPERLINKS IN WIKIPEDIA USING COLLECTIVE ATTENTION	1709
<i>Philip Thruesen ; Jaroslav Cechák ; Blandine Sezñec ; Roel Castalio ; Nattiya Kanhabua</i>	
AN OVERVIEW OF STUDIES ABOUT STUDENTS' PERFORMANCE ANALYSIS AND LEARNING ANALYTICS IN MOOCS	1719
<i>Ismail Duru ; Gulustan Dogan ; Banu Diri</i>	
SMART ONLINE VEHICLE TRACKING SYSTEM FOR SECURITY APPLICATIONS	1724
<i>Brahim Hnich ; Faisal R. Al-Osaimi ; Ata Sasmaz ; Özkan Sayin ; Amine Lamine ; Majid Alotaibi</i>	
AN OPTIMIZED FREQUENT PATTERN MINING ALGORITHM WITH MULTIPLE MINIMUM SUPPORTS	1734
<i>Hsiao-Wei Hu ; Hao-Chen Chang ; Wen-Shiu Lin</i>	
IMPROVING ITEM-BASED RECOMMENDATION ACCURACY WITH USER'S PREFERENCES ON APACHE MAHOUT	1742
<i>Ammar Jabakji ; Hasan Dag</i>	
CHANGE DETECTION AND CLASSIFICATION OF DIGITAL COLLECTIONS	1750
<i>Sampath Jayarathna ; Faryaneh Poursardar</i>	
A FEATURE SELECTION METHOD BASED ON LORENTZIAN METRIC	1760
<i>Yerzhan Kerimbekov ; Hasan Sakir Bilge</i>	
A SURVEY ON SEMANTIC WEB AND BIG DATA TECHNOLOGIES FOR SOCIAL NETWORK ANALYSIS	1768
<i>Sercan Kulcu ; Erdogan Dogdu ; A. Murat Ozbayoglu</i>	
TABLE CLASSIFICATION USING BOTH STRUCTURE AND CONTENT INFORMATION: A CASE STUDY OF FINANCIAL DOCUMENTS	1778
<i>Quanzhi Li ; Sameena Shah ; Rui Fang</i>	
PATIENT-RECORD LEVEL INTEGRATION OF DE-IDENTIFIED HEALTHCARE BIG DATABASES	1784
<i>Xiao Li ; Reza Sharifi Sedeh ; Liao Wang ; Yang Yang</i>	
A BAYESIAN PREDICTOR OF AIRLINE CLASS SEATS BASED ON MULTINOMIAL EVENT MODEL	1787
<i>Bingchuan Liu ; Yudong Tan ; Huimin Zhou</i>	
IDENTIFYING TROLLS AND DETERMINING TERROR AWARENESS LEVEL IN SOCIAL NETWORKS USING A SCALABLE FRAMEWORK	1792
<i>Busra Mutlu ; Merve Mutlu ; Kasim Oztoprak ; Erdogan Dogdu</i>	
DELAYRADAR: A MULTIVARIATE PREDICTIVE MODEL FOR TRANSIT SYSTEMS	1799
<i>Aparna Oruganti ; Fangzhou Sun ; Hiba Baroud ; Abhishek Dubey</i>	
A REAL-TIME AUTONOMOUS HIGHWAY ACCIDENT DETECTION MODEL BASED ON BIG DATA PROCESSING AND COMPUTATIONAL INTELLIGENCE	1807
<i>Murat Ozbayoglu ; Gokhan Kucukayan ; Erdogan Dogdu</i>	
SUBGROUP DISCOVERY ON BIG DATA: PRUNING THE SEARCH SPACE ON EXHAUSTIVE SEARCH ALGORITHMS	1814
<i>F. Padillo ; J. M. Luna ; S. Ventura</i>	
THE DIFFERENCE-OF-DATASETS FRAMEWORK: A STATISTICAL METHOD TO DISCOVER INSIGHT	1824
<i>Paul Raff ; Ze Jin</i>	
ONLINE TRAJECTORY SEGMENTATION AND SUMMARY WITH APPLICATIONS TO VISUALIZATION AND RETRIEVAL	1832
<i>Yehezkel S. Resheff</i>	
SKELETON DECOMPOSITION ANALYSIS FOR SUBSPACE CLUSTERING	1841
<i>Ali Sekmen ; Akram Aldroubi ; Ahmet Bugra Koku</i>	
AN EXTENDED IOT FRAMEWORK WITH SEMANTICS, BIG DATA, AND ANALYTICS	1849
<i>Omer Berat Sezer ; Erdogan Dogdu ; Murat Ozbayoglu ; Aras Onal</i>	

EVENT SEGMENTATION USING MAPREDUCE BASED BIG DATA CLUSTERING	1857
<i>M. Omair Shafiq</i>	
USER AND ENTITY BEHAVIOR ANALYTICS FOR ENTERPRISE SECURITY	1867
<i>Madhu Shashanka ; Min-Yi Shen ; Jisheng Wang</i>	
SWARM INTELLIGENCE (SI) BASED PROFILING AND SCHEDULING OF BIG DATA APPLICATIONS	1875
<i>Thamarai Selvi Somasundaram ; Kannan Govindarajan ; Vivekanandan Suresh Kumar</i>	
IMPROVING CLUSTERING EFFICIENCY BY SIMHASH-BASED K-MEANS ALGORITHM FOR BIG DATA ANALYTICS	1881
<i>Jenq-Haur Wang ; Jia-Zhi Lin</i>	
THE EFFECT OF PETS ON HAPPINESS: A DATA-DRIVEN APPROACH VIA LARGE-SCALE SOCIAL MEDIA	1889
<i>Yuchen Wu ; Jianbo Yuan ; Quanzeng You ; Jiebo Luo</i>	
INTELLIGENT AUTHORSHIP IDENTIFICATION WITH USING TURKISH NEWSPAPERS METADATA	1895
<i>Ozlem Yavanoglu</i>	
SOLVING COLD-START PROBLEM IN LARGE-SCALE RECOMMENDATION ENGINES: A DEEP LEARNING APPROACH	1901
<i>Jianbo Yuan ; Walid Shalaby ; Mohammed Korayem ; David Lin ; Khalifeh Aljadda ; Jiebo Luo</i>	
URBAN HUMAN MOBILITY DATA MINING: AN OVERVIEW	1911
<i>Kai Zhao ; Sasu Tarkoma ; Siyuan Liu ; Huy Vo</i>	
FINE-GRAINED MINING OF ILLICIT DRUG USE PATTERNS USING SOCIAL MULTIMEDIA DATA FROM INSTAGRAM	1921
<i>Yiheng Zhou ; Numair Sani ; Jiebo Luo</i>	
RESEARCH ON THE BIG DATA SYSTEM OF MASSIVE OPEN ONLINE COURSE	1931
<i>Zhenwei Du ; Haopeng Chen ; Jianwei Jiang</i>	
CLINICAL NAMED ENTITY RECOGNITION: CHALLENGES AND OPPORTUNITIES	1937
<i>Srinivasa Rao Kundeti ; J. Vijayananda ; Srikanth Mujjiga ; M. Kalyan</i>	
VERY FAST FREQUENT ITEMSET MINING: SIMPLICIAL COMPLEX METHODS (EXTENDED ABSTRACT)	1946
<i>Tsau-Young Lin</i>	
ONLINE ANOMALY DETECTION USING NON-PARAMETRIC TECHNIQUE FOR BIG DATA STREAMS IN CLOUD COLLABORATIVE ENVIRONMENT	1950
<i>G S Smrithy ; Sathyan Munirathinam ; Ramadoss Balakrishnan</i>	
A PROPOSAL OF A PRIVACY-PRESERVING QUESTIONNAIRE BY NON-DETERMINISTIC INFORMATION AND ITS ANALYSIS	1956
<i>Hiroshi Sakai ; Chenxi Liu ; Michinori Nakata ; Shusaku Tsumoto</i>	
PREDICTION OF INDIAN ELECTION USING SENTIMENT ANALYSIS ON HINDI TWITTER	1966
<i>Parul Sharma ; Teng-Sheng Moh</i>	
CONSTRUCTION OF CLINICAL PATHWAY FROM HISTORIES OF CLINICAL ACTIONS IN HOSPITAL INFORMATION SYSTEM	1972
<i>Shusaku Tsumoto ; Shoji Hirano ; Haruko Iwata</i>	
MINING PROCESS FOR IMPROVEMENT OF CLINICAL PROCESS QUALITY	1982
<i>Shusaku Tsumoto ; Shoji Hirano ; Haruko Iwata ; Norio Yoshimoto ; Tomohiro Kimura</i>	
MULTI-LAYER TEXT CLASSIFICATION WITH VOTING FOR CONSUMER REVIEWS	1991
<i>Yan Zhu ; Melody Moh ; Teng-Sheng Moh</i>	
SCEM: SMART & EFFECTIVE CROWD MANAGEMENT WITH A NOVEL SCHEME OF BIG DATA ANALYTICS	2000
<i>Shakti Awaghad</i>	
A SYSTEM AND ARCHITECTURE FOR REUSABLE ABSTRACTIONS OF MANUFACTURING PROCESSES	2004
<i>Alexander Brodsky ; Mohan Krishnamoorthy ; William Z. Bernstein ; M. Omar Nachawati</i>	
EVALUATION OF A PMML-BASED GPR SCORING ENGINE ON A CLOUD PLATFORM AND MICROCOMPUTER BOARD FOR SMART MANUFACTURING	2014
<i>Max Ferguson ; Kincho H. Law ; Raunak Bhinge ; David Dornfeld ; Jinkyoo Park ; Yung-Tsun Tina Lee</i>	
PREDICTING RARE FAILURE EVENTS USING CLASSIFICATION TREES ON LARGE SCALE MANUFACTURING DATA WITH COMPLEX INTERACTIONS	2024
<i>Jeff Hebert</i>	
USING BIG DATA TO ENHANCE THE BOSCH PRODUCTION LINE PERFORMANCE: A KAGGLE CHALLENGE	2029
<i>Ankita Mangal ; Nishant Kumar</i>	

BAYESIAN OPTIMIZATION FOR PREDICTING RARE INTERNAL FAILURES IN MANUFACTURING PROCESSES.....	2036
<i>Abhinav Maurya</i>	
MACHINE LEARNING, LINEAR AND BAYESIAN MODELS FOR LOGISTIC REGRESSION IN FAILURE DETECTION PROBLEMS	2046
<i>B. Pavlyshenko</i>	
CONVERGENCE AND DIVERGENCE IN ACADEMIC AND INDUSTRIAL INTERESTS ON IOT BASED MANUFACTURING.....	2051
<i>Srinivasan Radhakrishnan ; Sagar Kamarthi</i>	
COMPLEXITY-ENTROPY FEATURE PLANE FOR GEAR FAULT DETECTION.....	2057
<i>Srinivasan Radhakrishnan ; Sagar Kamarthi</i>	
CLOUD-BASED MACHINE LEARNING FOR PREDICTIVE ANALYTICS: TOOL WEAR PREDICTION IN MILLING	2062
<i>Dazhong Wu ; Connor Jennings ; Janis Terpenney ; Soundar Kumara</i>	
PREDICT FAILURES IN PRODUCTION LINES: A TWO-STAGE APPROACH WITH CLUSTERING AND SUPERVISED LEARNING	2070
<i>Darui Zhang ; Bin Xu ; Jasmine Wood</i>	
HOLISTIC DISASTER RECOVERY APPROACH FOR BIG DATA NOSQL WORKLOADS.....	2075
<i>Aharon Abadi ; Ashraf Haib ; Roie Melamed ; Alaa Nassar ; Aidan Shribman ; Hisham Yasin</i>	
DATA-DRIVEN CLOUD-BASED IT SERVICES PERFORMANCE FORECASTING	2081
<i>Genady Ya. Grabarnik ; Mauro Tortonesi ; Larisa Shwartz</i>	
ON-DEMAND DATA ANALYTICS IN HPC ENVIRONMENTS AT LEADERSHIP COMPUTING FACILITIES: CHALLENGES AND EXPERIENCES.....	2087
<i>John Harney ; Seung-Hwan Lim ; Sreenivas Sukumar ; Dale Stansberry ; Peter Xenopoulos</i>	
INTERCLOUD BROKERAGES BASED ON PLS METHOD FOR DEPLOYING INFRASTRUCTURES FOR BIG DATA ANALYTICS.....	2097
<i>Katsunori Miura ; Tazro Ohta ; Courtney Powell ; Masaharu Munetomo</i>	
MOTIVATING DYNAMIC FEATURES FOR RESOLUTION TIME ESTIMATION WITHIN IT OPERATIONS MANAGEMENT	2103
<i>Kayhan Moharreri ; Jayashree Ramanathan ; Rajiv Ramnath</i>	
IDENTIFYING PERFORMANCE BOTTLENECKS IN HIVE: USE OF PROCESSOR COUNTERS	2109
<i>Alexander C. Shulyak ; Lizy K. John</i>	
LEVERAGING LARGE SENSOR STREAMS FOR ROBUST CLOUD CONTROL	2115
<i>Alok Singh ; Eric Stephan ; Todd Elsethagen ; Matt Macduff ; Bibi Raju ; Malachi Schram ; Kerstin Kleese Van Dam ; Darren J Kerbyson ; Ilkay Altintas</i>	
FINE-GRAINED POWER ANALYSIS OF EMERGING GRAPH PROCESSING WORKLOADS FOR CLOUD OPERATIONS MANAGEMENT	2121
<i>Shuang Song ; Ximian Zheng ; Andreas Gerstlauer ; Lizy K. John</i>	
OPEN BIG DATA INFRASTRUCTURES TO EVERYONE	2127
<i>Konstantinos Tsakalozos ; Cory Johns ; Kevin Monroe ; Pete Vandergiessen ; Andrew McLeod ; Antonio Rosales</i>	
SPATIAL-CROWD: A BIG DATA FRAMEWORK FOR EFFICIENT DATA VISUALIZATION	2130
<i>Shahbaz Atta ; Bilal Sadiq ; Akhlaq Ahmad ; Sheikh Nasir Saeed ; Emad Felemban</i>	
MULTI-SCALAR ANALYSIS OF GEOSPATIAL AGRICULTURAL DATA FOR SUSTAINABILITY	2139
<i>Anne M. Denton ; Mostofa Ahsan ; David Franzen ; John Nowatzki</i>	
A FRAMEWORK FOR EVALUATING URBAN LAND USE MIX FROM CROWD-SOURCING DATA.....	2147
<i>Luciano Gervasoni ; Marti Bosch ; Serge Fenet ; Peter Sturm</i>	
CROWDSENSING AND ANALYZING MICRO-EVENT TWEETS FOR PUBLIC TRANSPORTATION INSIGHTS	2157
<i>Thong Hoang ; Pei Hua Cher ; Philips Kokoh Prasetyo ; Ee-Peng Lim</i>	
A STUDY FOR UNDERSTANDING OF TOURIST PERSON TRIP PATTERN BASED ON LOG DATA OF WI-FI ACCESS POINTS	2167
<i>Yu Ichifuji ; Yoshihide Matsuo ; Noriaki Koide ; Nobuhiro Akashi ; Yoshitaka Terai ; Toru Kobayashi</i>	
ESTIMATION OF NATIONAL TOURISM STATISTICS BASED ON WI-FI ASSOCIATION LOG DATA.....	2175
<i>Noriaki Koide ; Yu Ichifuji ; Hideki Yoshii ; Noboru Sonehara</i>	
PEER-TO-PEER MICROLENDING PLATFORMS: CHARACTERIZATION OF ONLINE TRAITS	2180
<i>Gaurav Paruthi ; Enrique Frias-Martinez ; Vanessa Frias-Martinez</i>	
NETWORK OPTIMIZATION OF FOOD FLOWS IN THE U.S.....	2190
<i>Caleb Robinson ; Arezoo Shirazi ; Mengmeng Liu ; Bistra Dilkina</i>	

MEASURING ACTIVITIES AND VALUES OF INDUSTRIAL CLUSTERS BASED ON JOB OPPORTUNITY DATA COLLECTED FROM AN INTERNET JAPANESE JOB MATCHING SITE	2199
<i>Aki-Hiro Sato ; Tsutomu Watanabe</i>	
SOLAR IRRADIANCE FORECASTING BY MACHINE LEARNING FOR SOLAR CAR RACES	2209
<i>Xiaoyan Shao ; Siyuan Lu ; Theodore G. Van Kessel ; Hendrik F. Hamann ; Leda Daehler ; Jeffrey Cwagenberg ; Alan Li</i>	
HOTEL PLAN POPULARITY FACTOR ANALYSIS OF HOTELS IN THE KEIHANSHIN REGION	2217
<i>Hiroshi Tsuda ; Masakazu Ando ; Yu Ichifuji</i>	
MIXED DATA AND CLASSIFICATION OF TRANSIT STOPS	2225
<i>Laura L. Tupper ; David S. Matteson ; John C. Handley</i>	
A SCALABLE AND COMPOSABLE MAP-REDUCE SYSTEM	2233
<i>Mahwish Arif ; Hans Vandierendonck ; Dimitrios S. Nikolopoulos ; Bronis R. De Supinski</i>	
A WORKLOAD AWARE MODEL OF COMPUTATIONAL RESOURCE SELECTION FOR BIG DATA APPLICATIONS	2243
<i>Amit Gupta ; Weijia Xu ; Natalia Ruiz-Juri ; Kenneth Perrine</i>	
EVALUATION OF K-MEANS DATA CLUSTERING ALGORITHM ON INTEL XEON PHI	2251
<i>Sunwoo Lee ; Wei-Keng Liao ; Ankit Agrawal ; Nikos Hardavellas ; Alok Choudhary</i>	
MATERIALS DISCOVERY: UNDERSTANDING POLYCRYSTALS FROM LARGE-SCALE ELECTRON PATTERNS	2261
<i>Ruoqian Liu ; Ankit Agrawal ; Wei-Keng Liao ; Alok Choudhary ; Marc De Graef</i>	
A BIG DATA PLATFORM INTEGRATING COMPRESSED LINEAR ALGEBRA WITH COLUMNAR DATABASES	2270
<i>Vishnu Gowda Harish ; Vinay Kumar Bingi ; John A. Miller</i>	
BUILDING A RESEARCH DATA SCIENCE PLATFORM FROM INDUSTRIAL MACHINES	2279
<i>Fang Cherry Liu ; Fu Shen ; Duen Horng Chau ; Neil Bright ; Mehmet Belgin</i>	
VISUALLY PROGRAMMING DATAFLOWS FOR DISTRIBUTED DATA ANALYTICS	2285
<i>Lauritz Thamsen ; Thomas Renner ; Marvin Byfeld ; Markus Paeschke ; Daniel Schroder ; Felix Böhm</i>	
BIG DATA ANALYTICS ON HPC ARCHITECTURES: PERFORMANCE AND COST	2295
<i>Peter Xenopoulos ; Jamison Daniel ; Michael Matheson ; Sreenivas Sukumar</i>	
SUPPORTING LARGE SCALE CONNECTED VEHICLE DATA ANALYSIS USING HIVE	2305
<i>Weijia Xu ; Natalia Ruiz Juri ; Amit Gupta ; Amanda Deering ; Chandra Bhat ; James Kuhr ; Jackson Archer</i>	
LEGION-BASED SCIENTIFIC DATA ANALYTICS ON HETEROGENEOUS PROCESSORS	2314
<i>Lina Yu ; Hongfeng Yu</i>	
ACCELERATING MATHEMATICAL KNOT SIMULATIONS WITH R ON THE WEB	2324
<i>Juan Lin ; Di Zhong ; Yiwen Zhong ; Hui Zhang</i>	
A GEOHYDROLOGIC DATA VISUALIZATION FRAMEWORK WITH AN EXTENDABLE USER INTERFACE DESIGN	2331
<i>Yanfu Zhou ; Jieting Wu ; Lina Yu ; Hongfeng Yu ; Zhenghong Tang</i>	
EFFICIENT PORTFOLIO ALLOCATION WITH SPARSE VOLATILITY ESTIMATION FOR HIGH-FREQUENCY FINANCIAL DATA	2341
<i>Jian Zou ; Chuqin Huang</i>	
DASK & NUMBA: SIMPLE LIBRARIES FOR OPTIMIZING SCIENTIFIC PYTHON CODE	2351
<i>James Crist</i>	
PINTERNET: A THEMATIC LABEL CURATION TOOL FOR LARGE IMAGE DATASETS	2353
<i>Ruoqian Liu ; Diana Palsetia ; Arindam Paul ; Reda Al-Bahrani ; Dipendra Jha ; Wei-Keng Liao ; Ankit Agrawal ; Alok Choudhary</i>	
IMPLEMENTING DICTIONARY LEARNING IN APACHE FLINK, OR: HOW I LEARNED TO RELAX AND LOVE ITERATIONS	2363
<i>Geoffrey Mon ; Milad Makkie ; Xiang Li ; Tianming Liu ; Shannon Quinn</i>	
MAKING MASSIVE COMPUTATIONAL EXPERIMENTS PAINLESS	2368
<i>Hatef Monajemi ; David L. Donoho ; Victoria Stodden</i>	
TOO BIG TO MAIL: ON THE WAY TO PUBLISH LARGE-SCALE MOBILE ANALYTICS DATA	2374
<i>Ella Peltonen ; Eemil Lagerspetz ; Petteri Nurmi ; Sasu Tarkoma</i>	
CONTENT-BASED RECOMMENDATION FOR PODCAST AUDIO-ITEMS USING NATURAL LANGUAGE PROCESSING TECHNIQUES	2378
<i>Zhou Xing ; Marzieh Parandehgheibi ; Fei Xiao ; Nilesh Kulkarni ; Chris Pouliot</i>	
A GLUE LANGUAGE FOR EVENT STREAM PROCESSING	2384
<i>Sylvain Hallé ; Sébastien Gaboury ; Raphaël Khoury</i>	

REAL-TIME PROCESSING OF PROTEOMICS DATA: THE INTERNET OF THINGS AND THE CONNECTED LABORATORY	2392
<i>Christopher Hillman ; Karen Petrie ; Andrew Cobley ; Mark Whitehorn</i>	
PREDICTING THE SHAPE AND PEAK TIME OF NEWS ARTICLE VIEWS	2400
<i>Yaser Keneshloo ; Shuguang Wang ; Eui-Hong Sam Han ; Naren Ramakrishnan</i>	
AN FPGA-BASED LOW-LATENCY NETWORK PROCESSING FOR SPARK STREAMING	2410
<i>Kohei Nakamura ; Ami Hayashi ; Hiroki Matsutani</i>	
HANDLING DELAYED LABELS IN TEMPORALLY EVOLVING DATA STREAMS	2416
<i>Joshua Plasse ; Niall Adams</i>	
A MULTI-LAYER SOFTWARE ARCHITECTURE FRAMEWORK FOR ADAPTIVE REAL-TIME ANALYTICS	2425
<i>Athena Vakali ; Paschalis Korosoglou ; Pavlos Daoglou</i>	
IMPLEMENTING TRAJECTORY DATA STREAM ANALYSIS IN PARALLEL	2431
<i>Yongyi Xian ; Chuanfei Xu ; Yan Liu</i>	
LANGUAGE INDEPENDENT BIG-DATA SYSTEM FOR THE PREDICTION OF USER LOCATION ON TWITTER	2437
<i>Jaime Alonso-Lorenzo ; Enrique Costa-Montenegro ; Milagros Fernández-Gavilanes</i>	
FORECASTING NIKE'S SALES USING FACEBOOK DATA	2447
<i>Linda Camilla Boldt ; Vinothan Vinayagamoorthy ; Florian Winder ; Melanie Schnittger ; Mats Ekran ; Raghava Rao Mukkamala ; Niels Buus Lassen ; Benjamin Flesch ; Abid Hussain ; Ravi Vratapu</i>	
FINDING INFORMATIVE COMMENTS FOR VIDEO VIEWING	2457
<i>Seungwoo Choi ; Aviv Segev</i>	
PREDICTION OF INFORMATION DIFFUSION IN SOCIAL NETWORKS USING DYNAMIC CARRYING CAPACITY	2466
<i>Anahita Davoudi ; Mainak Chatterjee</i>	
WHEN DO LUXURY CARS HIT THE ROAD? FINDINGS BY A BIG DATA APPROACH	2470
<i>Yang Feng ; Jiebo Luo</i>	
TWEET SENTIMENT AS PROXY FOR POLITICAL CAMPAIGN MOMENTUM	2475
<i>David Watts ; K. M. George ; T. K. Ashwin Kumar ; Zenia Arora</i>	
A NEW APPROACH TO BUILDING THE INTERINDUSTRY INPUT-OUTPUT TABLE USING BLOCK ESTIMATION TECHNIQUES	2485
<i>Ryohei Hisano</i>	
NOWCAST OF FIRM SALES USING POS DATA TOWARD STOCK MARKET STABILITY	2495
<i>Atushi Ishikawa ; Shouji Fujimoto ; Takayuki Mizuno</i>	
UNCOVERING INFORMATION FLOW AMONG USERS BY TIME-SERIES RETWEET DATA: WHO IS A FRIEND OF WHOM ON TWITTER?	2500
<i>Yuka Kamiko ; Mitsuo Yoshida ; Hirotada Ohashi ; Fujio Toriumi</i>	
QUANTIFYING MORAL FOUNDATIONS FROM VARIOUS TOPICS ON TWITTER CONVERSATIONS	2505
<i>Rishemjit Kaur ; Kazutoshi Sasahara</i>	
APPLICATION OF AN INTEGER-VALUED AUTOREGRESSIVE MODEL TO HIT PHENOMENA	2513
<i>Yasuko Kawahata ; Tamio Koyama</i>	
ANALYTICAL METHOD OF WEB USER BEHAVIOR USING HIDDEN MARKOV MODEL	2518
<i>Hirotaka Kawazu ; Fujio Toriumi ; Masanori Takano ; Kazuya Wada ; Ichiro Fukuda</i>	
LEVERAGING SOCIAL BIG DATA FOR PERFORMANCE EVALUATION OF E-COMMERCE WEBSITES	2525
<i>Eyad Makki ; Lin-Ching Chang</i>	
USER-GENERATED CONTENT CURATION WITH DEEP CONVOLUTIONAL NEURAL NETWORKS	2535
<i>Ruben Tous ; Otto Wust ; Mauro Gomez ; Jonatan Poveda ; Marc Elena ; Jordi Torres ; Mouna Makni ; Eduard Ayguadé</i>	
PRICING THE WOMAN CARD: GENDER POLITICS BETWEEN HILLARY CLINTON AND DONALD TRUMP	2541
<i>Yu Wang ; Yang Feng ; Jiebo Luo ; Xiyang Zhang</i>	
AUTOMATED CLASSIFICATION OF EXTREMIST TWITTER ACCOUNTS USING CONTENT-BASED AND NETWORK-BASED FEATURES	2545
<i>Udaniel Xie ; Jiejun Xu ; Tsai-Ching Lu</i>	
TOWARDS A HETEROGENEOUS, POLYSTORE-LIKE DATA ARCHITECTURE FOR THE US DEPARTMENT OF VETERAN AFFAIRS (VA) ENTERPRISE ANALYTICS	2550
<i>Edmon Begoli ; Derek Kistler ; Jack Bates</i>	

ANALYTICS-DRIVEN DATA INGESTION AND DERIVATION IN THE AWESOME POLYSTORE	2555
<i>Subhasis Dasgupta ; Kevin Coakley ; Amarnath Gupta</i>	
A SEMANTIC APPROACH TO POLYSTORES	2565
<i>E. Kharlamov ; T. Mailis ; K. Bereta ; D. Bilidas ; S. Brandt ; E. Jimenez-Ruiz ; S. Lamparter ; C. Neuenstadt ; O. Özgep ; A. Soyly ; C. Svingos ; G. Xiao ; D. Zheleznyakov ; D. Calvanese ; I. Horrocks ; M. Giese ; Y. Ioannidis ; Y. Kotidis ; R. Moller ; A. Waaler</i>	
BENCHMARKING POLYSTORES: THE CLOUDMDSQL EXPERIENCE	2574
<i>Boyan Kolev ; Raquel Pau ; Oleksandra Levchenko ; Patrick Valduries ; Ricardo Jiménez-Peris ; José Pereira</i>	
DIGREE: A MIDDLEWARE FOR A GRAPH DATABASES POLYSTORE	2580
<i>Vasilis Spyropoulos ; Christina Vasilakopoulou ; Yannis Kotidis</i>	
HOBBITS: HADOOP AND HIVE BASED INTERNET TRAFFIC ANALYSIS	2590
<i>Abdeltawab M. Hendawi ; Fatemah Alali ; Xiaoyu Wang ; Yunfei Guan ; Tianshu Zhou ; Xiao Liu ; Nada Basit ; John A. Stankovic</i>	
URBAN-NET: A NETWORK-BASED INFRASTRUCTURE MONITORING AND ANALYSIS SYSTEM FOR EMERGENCY MANAGEMENT AND PUBLIC SAFETY	2600
<i>Sangkeun Lee ; Liangzhe Chen ; Sisi Duan ; Supriya Chinthavali ; Mallikarjun Shankar ; B. Aditya Prakash</i>	
UNRAVELLING THE MYTH OF BIG DATA AND ARTIFICIAL INTELLIGENCE IN SUSTAINABLE NATURAL RESOURCE DEVELOPMENT	2610
<i>Gandhi Sivakumar ; Drew Johnson ; Rashida Hodge</i>	
BIG DATA COMPUTATION OF TAXI MOVEMENT IN NEW YORK CITY	2616
<i>Joya A. Deri ; Franz Franchetti ; José M. F. Moura</i>	
LINKED DATA VIEW METHODOLOGY AND APPLICATION TO BIM ALIGNMENT AND INTEROPERABILITY	2626
<i>Holly Ferguson ; Charles Vardeman ; Jarek Nabrzyski</i>	
THE SMART APPROACH TO COMPREHENSIVE QUALITY ASSESSMENT OF SITE-BASED SPATIAL-TEMPORAL DATA	2636
<i>Rafal A. Angrzyk ; Douglas E. Galarus</i>	
ADAPTING K-MEANS CLUSTERING TO IDENTIFY SPATIAL PATTERNS IN STORMS	2646
<i>Upa Gupta ; Kulsawasd Jitkajornwanich ; Ramez Elmasri ; Leonidas Fegaras</i>	
DETERMINING FEATURE EXTRACTORS FOR UNSUPERVISED LEARNING ON SATELLITE IMAGES	2655
<i>Behnam Hedayatnia ; Mehrdad Yazdani ; Mai Nguyen ; Jessica Block ; Ilkay Altintas</i>	
AN EXPERIMENTAL STUDY OF BIG SPATIAL DATA SYSTEMS	2664
<i>Andrew Hulbert ; Thomas Kunicki ; James N. Hughes ; Anthony D. Fox ; Christopher N. Eichelberger</i>	
IBM PAIRS CURATED BIG DATA SERVICE FOR ACCELERATED GEOSPATIAL DATA ANALYTICS AND DISCOVERY	2672
<i>Siyuan Lu ; Xiaoyan Shao ; Marcus Freitag ; Levente J. Klein ; Jason Renwick ; Fernando J. Marianno ; Conrad Albrecht ; Hendrik F. Hamann</i>	
A COMPARATIVE STUDY OF DUAL-TREE ALGORITHM IMPLEMENTATIONS FOR COMPUTING 2-BODY STATISTICS IN SPATIAL DATA	2676
<i>Chengcheng Mou ; Shaoping Chen ; Yi-Cheng Tu</i>	
TOWARDS A PROVENANCE-AWARE SPATIAL-TEMPORAL ARCHITECTURAL FRAMEWORK FOR MASSIVE DATA INTEGRATION AND ANALYSIS	2686
<i>Ivens Portugal ; Paulo Alencar ; Donald Cowan</i>	
USING PARALLEL HIERARCHICAL CLUSTERING TO ADDRESS SPATIAL BIG DATA CHALLENGES	2692
<i>Alan Woodley ; Ling-Xiang Tang ; Shlomo Geva ; Richi Nayak ; Timothy Chappell</i>	
BIG DATA DEVELOPMENT PLATFORM FOR ENGINEERING APPLICATIONS	2699
<i>Chien-Heng Wu ; Franco Lin ; Wen-Yi Chang ; Whey-Fone Tsai ; Hsi-Ching Lin ; Chao-Tung Yang</i>	
LARGE-SCALE SOLAR PANEL MAPPING FROM AERIAL IMAGES USING DEEP CONVOLUTIONAL NETWORKS	2703
<i>Jiangye Yuan ; Hsiu-Han Lexie Yang ; Olufemi A. Omitaomu ; Budhendra L. Bhaduri</i>	
SYMMETRIC REPOSITIONING OF BISECTING K-MEANS CENTERS FOR INCREASED REDUCTION OF DISTANCE CALCULATIONS FOR BIG DATA CLUSTERING	2709
<i>Yu Zhuang</i>	
EVALUATING MACHINE LEARNING ALGORITHMS FOR ANOMALY DETECTION IN CLOUDS	2716
<i>Anton Gulenko ; Marcel Wallschläger ; Florian Schmidt ; Odej Kao ; Feng Liu</i>	
PRELIMINARY BIG DATA IN A 5G TEST NETWORK	2722
<i>Teemu Kanstren ; Jussi Liikka ; Jukka Mäkelä ; Markus Luoto ; Jarmo Prokkola</i>	
QUICK MODEL FITTING USING A CLASSIFYING ENGINE	2728
<i>Yiming Kong ; Hui Zang ; Xiaoli Ma</i>	

SPARK-BASED RARE ASSOCIATION RULE MINING FOR BIG DATASETS	2734
<i>Ruilin Liu ; Kai Yang ; Yanjia Sun ; Tao Quan ; Jin Yang</i>	
WHAT: A BIG DATA APPROACH FOR ACCOUNTING OF MODERN WEB SERVICES	2740
<i>Martino Trevisan ; Idilio Drago ; Marco Mellia ; Han Hee Song ; Mario Baldi</i>	
BINARY: A FRAMEWORK FOR BIG DATA INTEGRATION FOR AD-HOC QUERYING	2746
<i>Azadeh Eftekhari ; Farhana Zulkernine ; Patrick Martin</i>	
CONTAINER-BASED VIRTUALIZATION FOR BYTE-ADDRESSABLE NVM DATA STORAGE	2754
<i>Ellis R. Giles</i>	
NOSQL SCHEMA EVOLUTION AND BIG DATA MIGRATION AT SCALE	2764
<i>Meike Klettke ; Uta Störl ; Manuel Shenavai ; Stefanie Scherzinger</i>	
SCHEDULING BIG DATA WORKFLOWS IN THE CLOUD UNDER BUDGET CONSTRAINTS	2775
<i>Aravind Mohan ; Mahdi Ebrahimi ; Shiyong Lu ; Alexander Kotov</i>	
BIG DATA AVAILABILITY: SELECTIVE PARTIAL CHECKPOINTING FOR IN-MEMORY DATABASE QUERIES	2785
<i>Daniel Playfair ; Amitabh Trehan ; Barry McLarnon ; Dimitrios S. Nikolopoulos</i>	
THE DIGITAL TRANSFORMATION AND SMART DATA ANALYTICS: AN OVERVIEW OF ENABLING DEVELOPMENTS AND APPLICATION AREAS	2795
<i>Nico Roedder ; David Dauer ; Kevin Laubis ; Paul Karaenke ; Christof Weinhardt</i>	
NON-DEEP CNN FOR MULTI-MODAL IMAGE CLASSIFICATION AND FEATURE LEARNING: AN AZURE-BASED MODEL	2803
<i>Sohini Roychowdhury ; Johnny Ren</i>	
TOWARDS AN INTEGRATED HEALTH RESEARCH PROCESS: A CLOUD-BASED APPROACH	2813
<i>Mathieu-P. Schapranow ; Matthias Uflacker ; Murat Sariyar ; Sebastian Semler ; Johannes Fichte ; Dietmar Schielke ; Kismet Ekinci ; Thomas Zahn</i>	
MODEL-DRIVEN DEPLOYMENT AND MANAGEMENT OF WORKFLOWS ON ANALYTICS FRAMEWORKS	2819
<i>Merlijn Sebrechts ; Sander Borny ; Thomas Vanhove ; Gregory Van Seghbroeck ; Tim Wauters ; Bruno Volckaert ; Filip De Turck</i>	
IS ELASTICITY OF SCALABLE DATABASES A MYTH?	2827
<i>Daniel Seybold ; Nicolas Wagner ; Benjamin Erb ; Jörg Domaschka</i>	
ANALYZING THE PERFORMANCE OF DATA REPLICATION AND DATA PARTITIONING IN THE CLOUD: THE BEOWULF APPROACH	2837
<i>Alexander Stiemer ; Ilir Fetai ; Heiko Schuldt</i>	
UNDERSTANDING PERFORMANCE INTERFERENCE IN MULTI-TENANT CLOUD DATABASES AND WEB APPLICATIONS	2847
<i>Miguel G. Xavier ; Kassiano J. Matteussi ; Fabian Lorenzo ; Cesar A. F. De Rose</i>	
EVALUATION-DRIVEN RESEARCH IN DATA SCIENCE: LEVERAGING CROSS-FIELD METHODOLOGIES	2853
<i>Bonnie J. Dorr ; Peter C. Fontana ; Craig S. Greenberg ; Marion Le Bras ; Mark Przybocki</i>	
BAD BIG DATA SCIENCE	2863
<i>Frank S. Haug</i>	
BIG DATA TEAM PROCESS METHODOLOGIES: A LITERATURE REVIEW AND THE IDENTIFICATION OF KEY FACTORS FOR A PROJECT'S SUCCESS	2872
<i>Jeffrey S. Saltz ; Ivan Shamshurin</i>	
PROGRESSION ANALYSIS OF SIGNALS: EXTENDING CRISP-DM TO STREAM ANALYTICS	2880
<i>Pankush Kalgotra ; Ramesh Sharda</i>	
SOFTWARE ENGINEERING FOR BIG DATA PROJECTS: DOMAINS, METHODOLOGIES AND GAPS	2886
<i>Vijay Dipti Kumar ; Paulo Alencar</i>	
NOT ALL SOFTWARE ENGINEERS CAN BECOME GOOD DATA ENGINEERS	2896
<i>Jeffrey S. Saltz ; Sibel Yilmazel ; Ozgur Yilmazel</i>	
A HACKING TOOLSET FOR BIG TABULAR FILES (CODENAMES: BIN4TSV, KABUTOMUSHI)	2902
<i>Toshiyuki Shimono</i>	
DISTRIBUTED AND CLOUD-BASED MULTI-MODEL ANALYTICS EXPERIMENTS ON LARGE VOLUMES OF CLIMATE CHANGE DATA IN THE EARTH SYSTEM GRID FEDERATION ECO-SYSTEM	2911
<i>S. Fiore ; M. Plóciennik ; C. Doutriaux ; C. Palazzo ; J. Boute ; T. Zok ; D. Elia ; M. Owsiak ; A. D'Anca ; Z. Shaheen ; R. Bruno ; M. Fargetta ; M. Caballer ; G. Moltó ; I. Blanquer ; R. Barbera ; M. David ; G. Donvito ; D. N. Williams ; V. Anantharaj ; D. Salomoni ; G. Aloisio</i>	

MODELING MARTIAN THERMAL INERTIA IN A DISTRIBUTED MEMORY HIGH PERFORMANCE COMPUTING ENVIRONMENT	2919
<i>Jason Laura ; Robin L. Fergason</i>	
WHERE BIG DATA MEETS LINKED DATA: APPLYING STANDARD DATA MODELS TO ENVIRONMENTAL DATA STREAMS.....	2929
<i>Adam Leadbetter ; Damian Smyth ; Robert Fuller ; Eoin O'Grady ; Adam Shepherd</i>	
THREE-DIMENSIONAL SPATIAL JOIN COUNT EXPLOITING CPU OPTIMIZED STR R-TREE	2938
<i>Ryuya Mitsuhashi ; Hideyuki Kawashima ; Takahiro Nishimichi ; Osamu Tatebe</i>	
IMPLEMENTING CONNECTED COMPONENT LABELING AS A USER DEFINED OPERATOR FOR SCIDB	2948
<i>Amidu Oloso ; Kwo-Sen Kuo ; Thomas Clune ; Paul Brown ; Alex Poliakov ; Hongfeng Yu</i>	
A NEW PARALLEL PYTHON TOOL FOR THE STANDARDIZATION OF EARTH SYSTEM MODEL DATA.....	2953
<i>Kevin Paul ; Sheri Mickelson ; John M. Dennis</i>	
USING CLOUD BURSTING TO COUNT TREES AND SHRUBS IN SUB-SAHARAN AFRICA	2960
<i>Michael Requa ; Garrison Vaughan ; John David ; Ben Cotton</i>	
SCISPARK: HIGHLY INTERACTIVE IN-MEMORY SCIENCE DATA ANALYTICS	2964
<i>Brian Wilson ; Rahul Palamuttam ; Kim Whitehall ; Chris Mattmann ; Alex Goodman ; Mazyar Boustani ; Sujen Shah ; Paul Zimdars ; Paul Ramirez</i>	
VISUALIZATION AND DIAGNOSIS OF EARTH SCIENCE DATA THROUGH HADOOP AND SPARK	2974
<i>Shujia Zhou ; Xiaowen Li ; Toshihisa Matsui ; Weikuo Tao</i>	
PERSISTING IN-MEMORY DATABASES USING SCM	2981
<i>Ellis Giles ; Kshitij Doshi ; Peter Varman</i>	
SS-DEDUP: A HIGH THROUGHPUT STATEFUL DATA ROUTING ALGORITHM FOR CLUSTER DEDUPLICATION SYSTEM	2991
<i>Zhihao Huang ; Hui Li ; Xin Li ; Wei He</i>	
ESTORE: AN EFFECTIVE OPTIMIZED DATA PLACEMENT STRUCTURE FOR HIVE	2996
<i>Xin Li ; Hui Li ; Zhihao Huang ; Bing Zhu ; Jiawei Cai</i>	
TOWARDS OPTIMIZING LARGE-SCALE DATA TRANSFERS WITH END-TO-END INTEGRITY VERIFICATION.....	3002
<i>Si Liu ; Eun-Sung Jung ; Rajkumar Kettimuthu ; Xian-He Sun ; Michael Papka</i>	
COLOC: DISTRIBUTED DATA AND CONTAINER COLOCATION FOR DATA-INTENSIVE APPLICATIONS.....	3008
<i>Thomas Renner ; Lauritz Thamsen ; Odej Kao</i>	
LINKED DATA PLATFORM FOR BUILDING CLOUD-BASED SMART APPLICATIONS AND CONNECTING API ACCESS POINTS WITH DATA DISCOVERY TECHNIQUES	3016
<i>Holly Ferguson ; Charles Vardeman ; Jarek Nabrzyski</i>	
METASTORE: A METADATA FRAMEWORK FOR SCIENTIFIC DATA REPOSITORIES.....	3026
<i>Ajinkya Prabhune ; Hasebullah Ansari ; Anil Keshav ; Rainer Stotzka ; Michael Gertz ; Jürgen Hesser</i>	
AUTOMATED SCHEMA EXTRACTION FOR PID INFORMATION TYPES	3036
<i>Ulrich Schwardmann</i>	
FACILITATING REPRODUCIBLE RESEARCH BY INVESTIGATING COMPUTATIONAL METADATA.....	3045
<i>Priyaa Thavasimani ; Paolo Missier</i>	
CONSTELLATION: A SCIENCE GRAPH NETWORK FOR SCALABLE DATA AND KNOWLEDGE DISCOVERY IN EXTREME-SCALE SCIENTIFIC COLLABORATIONS.....	3052
<i>Sudharshan S. Vazhkudai ; John Harney ; Raghu Gunasekaran ; Dale Stansberry ; Seung-Hwan Lim ; Tom Barron ; Andrew Nash ; Arvind Ramanathan</i>	
DETECTING SPAMMERS ON SOCIAL NETWORKS BASED ON A HYBRID MODEL.....	3062
<i>Guangxia Xu ; Jin Qi ; Deling Huang ; Mahmoud Daneshmand</i>	
BANDWIDTH PROVISION STRATEGIES FOR RELIABLE DATA MOVEMENTS IN DEDICATED NETWORKS.....	3069
<i>Liudong Zuo ; Michelle Mengxia Zhu</i>	
INVESTIGATION OF FORECASTING METHODS FOR THE HOURLY SPOT PRICE OF THE DAY-AHEAD ELECTRIC POWER MARKETS	3079
<i>Radhakrishnan Angamuthu Chinnathambi ; Prakash Ranganathan</i>	
LEVERAGING USER EXPERTISE IN COLLABORATIVE SYSTEMS FOR ANNOTATING ENERGY DATASETS	3087
<i>Hông-ân Cao ; Felix Rauchenstein ; Tri Kurniawan Wijaya ; Karl Aberer ; Nuno Nunes</i>	

TEMPORAL ASSOCIATION RULES FOR ELECTRICAL ACTIVITY DETECTION IN RESIDENTIAL HOMES	3097
<i>Hông-ân Cao ; Tri Kurniawan Wijaya ; Karl Aberer ; Nuno Nunes</i>	
LEVERAGING CLOUD COMPUTING TO CONVERT THE NON-INTRUSIVE LOAD MONITOR INTO A POWERFUL FRAMEWORK FOR GRID-RESPONSIVE BUILDINGS	3107
<i>Saman Mostafavi ; Benjamin Futrell ; John Troxler ; Robert W. Cox</i>	
BIG DATA, BETTER ENERGY MANAGEMENT AND CONTROL DECISIONS FOR DISTRIBUTION SYSTEMS IN SMART GRID	3115
<i>Shady S. Refaat ; Haitham Abu-Rub ; Amira Mohamed</i>	
DETECTING NON-TECHNICAL ENERGY LOSSES THROUGH STRUCTURAL PERIODIC PATTERNS IN AMI DATA	3121
<i>Viktor Botev ; Magnus Almgren ; Vincenzo Gulisano ; Olaf Landsiedel ; Marina Papatriantafilou ; Joris Van Rooij</i>	
LOSSLESS COMPRESSION OF HIGH-FREQUENCY VOLTAGE AND CURRENT DATA IN SMART GRIDS	3131
<i>Andreas Unterweger ; Dominik Engel</i>	
INDEXING SPATIOTEMPORAL RELATIONS IN SOLAR EVENT DATASETS	3140
<i>Berkay Aydın ; Ahmet Kucuk ; Rafal A. Angryk</i>	
SPATIO-TEMPORAL INTERPOLATION METHODS FOR SOLAR EVENTS METADATA	3149
<i>Soukaina Filali Boubrahimi ; Berkay Aydın ; Dustin Kempton ; Rafal Angryk</i>	
PROCESSING AND MANAGING THE KEPLER MISSION'S TREASURE TROVE OF STELLAR AND EXOPLANET DATA	3158
<i>Jon M. Jenkins</i>	
DESCRIBING SOLAR IMAGES WITH SPARSE CODING FOR SIMILARITY SEARCH	3168
<i>Dustin J. Kempton ; Michael A. Schuh ; Rafal A. Angryk</i>	
A DATA-DRIVEN ANALYSIS OF INTERPLANETARY CORONAL MASS EJECTA AND MAGNETIC FLUX ROPES	3177
<i>Ruizhe Ma ; Rafal Angryk ; Pete Riley</i>	
RUNNING SCIENTIFIC ALGORITHMS AS ARRAY DATABASE OPERATORS: BRINGING THE PROCESSING POWER TO THE DATA	3187
<i>Simon Marcin ; André Csillaghy</i>	
THE BEST OF BOTH WORLDS: USING AUTOMATIC DETECTION AND LIMITED HUMAN SUPERVISION TO CREATE A HOMOGENOUS MAGNETIC CATALOG SPANNING FOUR SOLAR CYCLES	3194
<i>A. Muñoz-Jaramillo ; Z. A. Werginz ; J. P. Vargas-Acosta ; M. D. Deluca ; J. C. Windmueller ; J. Zhang ; D. W. Longcope ; D. A. Lamb ; C. E. Deforest ; S. Vargas-Domínguez ; J. W. Harvey ; P. C. H. Martens</i>	
AN INPUT CATALOG AND TARGET SELECTION FOR THE TRANSITING EXOPLANET SURVEY SATELLITE	3204
<i>Ryan J. Oelkers ; Keivan G. Stassun ; Joshua Pepper ; Nathan De Lee ; Martin Paegert</i>	
METHOD FOR ESTIMATING CYCLE LENGTHS FROM MULTIDIMENSIONAL TIME SERIES: TEST CASES AND APPLICATION TO A MASSIVE "IN SILICO" DATASET	3214
<i>N. Olsper ; M. J. Käpylä ; J. Pelt</i>	
OPENING UP DARK DIGITAL ARCHIVES THROUGH THE USE OF ANALYTICS TO IDENTIFY SENSITIVE CONTENT	3224
<i>Bennett B. Borden ; Jason R. Baron</i>	
MINING AND ANALYSING ONE BILLION REQUESTS TO LINGUISTIC SERVICES	3230
<i>Marco Büchler ; Greta Franzini ; Emily Franzini ; Thomas Eckart</i>	
MIND THE EXPLANATORY GAP: QUALITY FROM QUANTITY	3240
<i>Jenny Bunn</i>	
EXPLORING ARCHIVES WITH PROBABILISTIC MODELS: TOPIC MODELLING FOR THE VALORISATION OF DIGITISED ARCHIVES OF THE EUROPEAN COMMISSION	3245
<i>Simon Hengchen ; Mathias Coeckelbergs ; Seth Van Hooland ; Ruben Verborgh ; Thomas Steiner</i>	
UNDERSTANDING COMPUTATIONAL WEB ARCHIVES RESEARCH METHODS USING RESEARCH OBJECTS	3250
<i>Emily Maemura ; Christoph Becker ; Ian Milligan</i>	
TRACES THROUGH TIME: A PROBABILISTIC APPROACH TO CONNECTED ARCHIVAL DATA	3260
<i>Sonia Ranade</i>	
COMPUTATIONAL PROVENANCE: DATAONE AND IMPLICATIONS FOR CULTURAL HERITAGE INSTITUTIONS	3266
<i>Robert J. Sandusky</i>	

APPRAISING DIGITAL ARCHIVES WITH ARCHIVEMATICA	3272
<i>Michael Shallcross</i>	
BREAKING DOWN THE INVISIBLE WALL TO ENRICH ARCHIVAL SCIENCE AND PRACTICE	3277
<i>Kenneth Thibodeau</i>	
CONTENT-BASED COMPARISON FOR COLLECTIONS IDENTIFICATION	3283
<i>Weijia Xu ; Ruizhu Huang ; Maria Esteva ; Jawon Song ; Ramona Walls</i>	
DEEP TOPOLOGY CLASSIFICATION: A NEW APPROACH FOR MASSIVE GRAPH CLASSIFICATION	3290
<i>Stephen Bonner ; John Brennan ; Georgios Theodoropoulos ; Ibad Kureshi ; Andrew Stephen McGough</i>	
GFP-X: A PARALLEL APPROACH TO MASSIVE GRAPH COMPARISON USING SPARK	3298
<i>Stephen Bonner ; John Brennan ; Georgios Theodoropoulos ; Ibad Kureshi ; Andrew Stephen McGough</i>	
FAST DISTRIBUTED K-NN GRAPH UPDATE	3308
<i>Thibault Debatty ; Fabio Pulvirenti ; Pietro Michiardi ; Wim Mees</i>	
AN INCREMENTAL LOCAL-FIRST COMMUNITY DETECTION METHOD FOR DYNAMIC GRAPHS	3318
<i>Hiroki Kanezashi ; Toyotaro Suzumura</i>	
MASSIVE GRAPH PROCESSING ON NANOCOMPUTERS	3326
<i>Bryan Rainey ; David F. Gleich</i>	
GRAPHFLOW: WORKFLOW-BASED BIG GRAPH PROCESSING	3336
<i>Sara Riaz ; Boyana Norris</i>	
ON THE HYPERBOLICITY OF LARGE-SCALE NETWORKS AND ITS ESTIMATION	3344
<i>W. Sean Kennedy ; Iraj Saniee ; Onuttom Narayan</i>	
PARALLEL GRAPH MINING WITH DYNAMIC LOAD BALANCING	3352
<i>Nilothpal Talukder ; Mohammed J. Zaki</i>	
DISTRIBUTED EXACT SUBGRAPH MATCHING IN SMALL DIAMETER DYNAMIC GRAPHS	3360
<i>Charith Wickramaarachchi ; Rajgopal Kannan ; Charalampos Chelms ; Viktor K. Prasanna</i>	
FAST REACHABILITY QUERY COMPUTATION ON BIG ATTRIBUTED GRAPHS	3370
<i>Duncan Yung ; Shi-Kuo Chang</i>	
DRUG TARGET PATH DISCOVERY ON SEMANTIC BIOMEDICAL BIG DATA	3381
<i>Fang Du ; Ting Li ; Yingjie Shi ; Lijuan Song ; Xiaojun Gu</i>	
A FRAMEWORK TO PREDICT OUTCOME FOR CANCER PATIENTS USING DATA FROM A NURSING EHR	3387
<i>Muhammad K Lodhi ; Rashid Ansari ; Yingwei Yao ; Gail M. Keenan ; Diana J. Wilkie ; Ashfaq Khokhar</i>	
DISTRIBUTED RANK-1 DICTIONARY LEARNING: TOWARDS FAST AND SCALABLE SOLUTIONS FOR FMRI BIG DATA ANALYTICS	3396
<i>Milad Makkie ; Xiang Li ; Tianming Liu ; Shannon Quinn ; Binbin Lin ; Jieping Ye</i>	
MORTALITY PREDICTION OF ICU PATIENTS USING LAB TEST DATA BY FEATURE VECTOR COMPACTION & CLASSIFICATION	3404
<i>Mohammad M. Masud ; Abdel Rahman Al Harahsheh</i>	
ITERATIVE UNIFIED CLUSTERING IN BIG DATA	3412
<i>Vasundhara Misal ; Vandana P. Janeja ; Sai. C. Pallaprolu ; Yelena Yesha ; Raghu Chintalapati</i>	
APPLICATION OF BIG DATA ANALYTICS FOR AUTOMATED ESTIMATION OF CT IMAGE QUALITY	3422
<i>Maitham D Naeemi ; Johnny Ren ; Nathan Hollcroft ; Adam M Alessio ; Sohini Roychowdhury</i>	
WEARABLE SENSOR BASED HUMAN POSTURE RECOGNITION	3432
<i>Jianwu Wang ; Zhichuan Huang ; Wenbin Zhang ; Ankita Patil ; Ketan Patil ; Ting Zhu ; Eric J Shiroma ; Mitchell A Schepps ; Tamara B Harris</i>	
SIMPLE AND EFFECTIVE PRE-PROCESSING FOR AUTOMATED MELANOMA DISCRIMINATION BASED ON CYTOLOGICAL FINDINGS	3439
<i>Takuya Yoshida ; M. Emre Celebi ; Gerald Schaefer ; Hitoshi Iyatomi</i>	
BIG DATA APPROACH IN HEALTHCARE USED FOR INTELLIGENT DESIGN — SOFTWARE AS A SERVICE	3443
<i>Weider D. Yu ; Jaspal Singh Gill ; Maulin Dalal ; Piyush Jha ; Sajan Shah</i>	
INTERACTIVE PERSONALIZED INTERESTING PATTERN DISCOVERY	3450
<i>Mansurul A Bhuiyan ; Mohammad Al Hasan</i>	
ANDROID MALWARE DETECTION WITH WEAK GROUND TRUTH DATA	3457
<i>Jordan Deloach ; Doina Caragea ; Xinming Ou</i>	
PROBABILISTIC PARALLELISATION OF BLOCKING NON-MATCHED RECORDS FOR BIG DATA	3465
<i>Chenxiao Dou ; Daniel Sun ; Yi-Cheng Chen ; Guoqiang Li ; Jianquan Liu</i>	

UNIVERSAL DATA DISCOVERY USING ATYPICALITY	3474
<i>Anders Host-Madsen ; Elyas Sabeti ; Chad Walton ; Su Jun Lim</i>	
A MARKOV CHAIN COLLABORATIVE FILTERING MODEL FOR COURSE ENROLLMENT RECOMMENDATIONS	3484
<i>Elham S. Khorasani ; Zhao Zhenge ; John Champaign</i>	
PREDICTING TRAFFIC OF ONLINE ADVERTISING IN REAL-TIME BIDDING SYSTEMS FROM PERSPECTIVE OF DEMAND-SIDE PLATFORMS	3491
<i>Hsu-Chao Lai ; Wen-Yueh Shih ; Jiun-Long Huang ; Yi-Cheng Chen</i>	
LEVERAGING CLOUD DATA TO MITIGATE USER EXPERIENCE FROM ‘BREAKING BAD’	3499
<i>Nicholas A. James ; Arun Kejariwal ; David S. Matteson</i>	
TOPIC MODELING FOR MANAGEMENT SCIENCES: A NETWORK-BASED APPROACH	3509
<i>Max Menenberg ; Surya Pathak ; Hari P. Udyapuram ; Srinagesh Gavirneni ; Sohini Roychowdhury</i>	
THE TECHNICAL HASHTAG IN TWITTER DATA: A HADOOP EXPERIENCE	3519
<i>Izabela Moise</i>	
USING SEMANTIC-BASED APPROACH TO MANAGE PERSPECTIVES OF PROCESS MINING: APPLICATION ON IMPROVING LEARNING PROCESS DOMAIN DATA	3529
<i>Okoye Kingsley ; Abdel-Rahman H. Tawil ; Usman Naeem ; Syed Islam ; Elyes Lamine</i>	
LABEL PROPAGATION IN BIG DATA TO DETECT REMOTE ACCESS TROJANS	3539
<i>Sai C. Pallaprolu ; Josephine M. Namayanja ; Vandana P. Janeja ; C. T. Sai Adithya</i>	
A NOVEL BIG-DATA PROCESSING FRAMEWORK FOR HEALTHCARE APPLICATIONS: BIG-DATA-HEALTHCARE-IN-A-BOX	3548
<i>Fuad Rahman ; Marvin Slepian ; Ari Mitra</i>	
AN EFFICIENT PARALLEL TOPIC-SENSITIVE EXPERT FINDING ALGORITHM USING SPARK	3556
<i>Yao-Ming Yang ; Chang-Dong Wang ; Jian-Huang Lai</i>	
EXPLORING THE UTILIZATION OF PLACES THROUGH A SCALABLE “ACTIVITIES IN PLACES” ANALYSIS MECHANISM	3563
<i>Linlin You ; Bige Tuncer</i>	
ROBUST K-SUBSPACES RECOVERY WITH COMBINATORIAL INITIALIZATION	3573
<i>Jun He ; Yue Zhang ; Jiye Wang ; Nan Zeng ; Hanyong Hao</i>	
TSMAP3D: BROWSER VISUALIZATION OF HIGH DIMENSIONAL TIME SERIES DATA	3583
<i>Supun Kamburugamuve ; Pulasthi Wickramasinghe ; Saliya Ekanayake ; Chathuri Wimalasena ; Milinda Pathirage ; Geoffrey Fox</i>	
ON THE THEORY AND PRACTICE OF HIGH-DIMENSIONAL DATA INDEXING WITH IDISTANCE	3593
<i>Michael A. Schuh ; Rafal A. Angryk</i>	
“INFLUENCE SKETCHING”: FINDING INFLUENTIAL SAMPLES IN LARGE-SCALE REGRESSIONS	3601
<i>Mike Wojnowicz ; Ben Cruz ; Xuan Zhao ; Brian Wallace ; Matt Wolff ; Jay Luan ; Caleb Crable</i>	
MINIMUM DENSITY HYPERPLANES IN THE FEATURE SPACE	3613
<i>Katie R. Yates ; Nicos G. Pavlidis</i>	
STRUCTURE PRESERVING DIMENSION REDUCTION WITH 2D IMAGES AS PREDICTORS	3619
<i>Bo Zhang ; Liwei Wang</i>	
MEMORY ACCESS PATTERN BASED INSIDER THREAT DETECTION IN BIG DATA SYSTEMS	3625
<i>Santosh Aditham ; Nagarajan Ranganathan ; Srinivas Katkooari</i>	
AUTOMATED BIG SECURITY TEXT PRUNING AND CLASSIFICATION	3629
<i>Khudran Alzhrani ; Ethan M. Rudd ; C. Edward Chow ; Terrance E. Boulton</i>	
BIG DATA ANALYTICS AS-A-SERVICE: ISSUES AND CHALLENGES	3638
<i>Claudio A. Ardagna ; Paolo Ceravolo ; Ernesto Damiani</i>	
DATA PRIVACY FOR IOT SYSTEMS: CONCEPTS, APPROACHES, AND RESEARCH DIRECTIONS	3645
<i>Elisa Bertino</i>	
TOWARDS AN EFFECTIVE AND EFFICIENT MALWARE DETECTION SYSTEM	3648
<i>Chia Tien Dan Lo ; Ordonez Pablo ; Cepeda Mora Carlos</i>	
PRIVATE DATABASES ON THE CLOUD: MODELS, ISSUES AND RESEARCH PERSPECTIVES	3656
<i>Alfredo Cuzzocrea ; Carlo Mastroianni ; Giorgio Mario Grasso</i>	
CONCISE ESSENCE-PRESERVING BIG DATA REPRESENTATION	3662
<i>Philip Derbeko ; Shlomi Dolev ; Ehud Gudes ; Jeffrey D. Ullman</i>	
TRUSTED CLOUD SQL DBS WITH ON-THE-FLY AES DECRYPTION/ENCRYPTION	3666
<i>Sushil Jajodia ; Witold Litwin ; Thomas Schwarz</i>	

AN ENTROPY-BASED ANALYTIC MODEL FOR THE PRIVACY-PRESERVING IN OPEN DATA.....	3676
<i>Soo-Hyung Kim ; Changwook Jung ; Yoon-Joon Lee</i>	
PHISHING DETECTION BASED ON NEWLY REGISTERED DOMAINS.....	3685
<i>Xueni Li ; Guanggang Geng ; Zhiwei Yan ; Yong Chen ; Xiaodong Lee</i>	
SECURITY AND PRIVACY FOR BIG DATA: A SYSTEMATIC LITERATURE REVIEW	3693
<i>Boel Nelson ; Tomas Olovsson</i>	
PHISHING THROUGH SOCIAL BOTS ON TWITTER.....	3703
<i>Mohammad Shafahi ; Leon Kempers ; Hamideh Afsarmanesh</i>	
REVERSE ENGINEERING SMART CARD MALWARE USING SIDE CHANNEL ANALYSIS WITH MACHINE LEARNING TECHNIQUES.....	3713
<i>Hippolyte Djonon Tsague ; Bheki Twala</i>	
S3C: AN ARCHITECTURE FOR SPACE-EFFICIENT SEMANTIC SEARCH OVER ENCRYPTED DATA IN THE CLOUD	3722
<i>Jason Woodworth ; Mohsen Amini Salehi ; Vijay Raghavan</i>	
A SYSTEMS APPROACH TO BIG DATA TECHNOLOGY APPLIED TO SUPPLY CHAIN.....	3732
<i>Tomohiro Fukui</i>	
OPTIMIZING PERFORMANCE OF SENTIMENT ANALYSIS THROUGH DESIGN OF EXPERIMENTS.....	3737
<i>Gary S. W. Goh ; Andy J. L. Ang ; Allan N. S. Zhang</i>	
ANALYSIS FOR SUPPLY HUB IN INDUSTRIAL CLUSTER: CLASSIC VS. NEW PERSPECTIVE	3743
<i>Vahid Kayvanfar ; S. M. Moattar Husseini ; Behrooz Karimi ; Mohsen S. Sajadieh ; Tan Wen Jun</i>	
A DEA APPROACH FOR SUPPLIER SELECTION WITH AHP AND RISK CONSIDERATION.....	3749
<i>Jasmine J. Lim ; Allan N. Zhang</i>	
DEEP LEARNING IN THE AUTOMOTIVE INDUSTRY: APPLICATIONS AND TOOLS.....	3759
<i>Andre Luckow ; Matthew Cook ; Nathan Ashcraft ; Edwin Weill ; Emil Djerekarov ; Bennie Vorster</i>	
THE BAYESIAN ESTIMATORS OF POLYTOMOUS ITEM RESPONSE THEORY MODELS WITH APPROXIMATED CONDITIONAL LIKELIHOOD AND THEIR MATHEMATICAL OPTIMALITIES	3769
<i>Kazumasa Mori ; Takuya Ohmori</i>	
DATA BLENDING IN MANUFACTURING AND SUPPLY CHAINS.....	3773
<i>B. Y. Ong ; Rong Wen ; Allan N. Zhang</i>	
ADAPTIVE RESILIENT STRATEGIES FOR SUPPLY CHAIN NETWORKS	3779
<i>Wen Jun Tan ; Wentong Cai ; Zhengping Li</i>	
PREDICTION OF REGIONAL GOODS DEMAND INCORPORATING THE EFFECT OF WEATHER	3785
<i>Takuya Watanabe ; Hiroaki Muroi ; Motoki Naruke ; Kyoto Yono ; Gen Kobayashi ; Masanori Yamasaki</i>	
WEIGHTED CLUSTERING OF SPATIAL PATTERN FOR OPTIMAL LOGISTICS HUB DEPLOYMENT.....	3792
<i>Rong Wen ; Wenjing Yan ; Allan N. Zhang</i>	
VESSEL MOVEMENT ANALYSIS AND PATTERN DISCOVERY USING DENSITY-BASED CLUSTERING APPROACH	3798
<i>Wenjing Yan ; Rong Wen ; Allan N. Zhang ; Dazhi Yang</i>	
SPATIAL DATA DIMENSION REDUCTION USING QUADTREE: A CASE STUDY ON SATELLITE-DERIVED SOLAR RADIATION	3807
<i>Dazhi Yang ; Gary S. W. Goh ; Siwei Jiang ; Allan N. Zhang</i>	
FORECAST UPC-LEVEL FMCG DEMAND, PART III: GROUPED RECONCILIATION.....	3813
<i>Dazhi Yang ; Gary S. W. Goh ; Siwei Jiang ; Allan N. Zhang</i>	
WORD EMBEDDINGS FOR ARABIC SENTIMENT ANALYSIS.....	3820
<i>A. Aziz Altowayan ; Lixin Tao</i>	
GIVING VOICE TO OFFICE CUSTOMERS: BEST PRACTICES IN HOW OFFICE HANDLES VERBATIM TEXT FEEDBACK	3826
<i>Michael Bentley ; Soumya Batra</i>	
UNLOCK BIG DATA EMOTIONS: WEIGHTED WORD EMBEDDINGS FOR SENTIMENT CLASSIFICATION.....	3833
<i>Xiangfeng Dai ; Bob Prout</i>	
BIG SOCIAL DATA ANALYTICS OF CHANGES IN CONSUMER BEHAVIOUR AND OPINION OF A TV BROADCASTER.....	3839
<i>Anna Hennig ; Anne-Sofie Åmodt ; Henrik Hernes ; Helene Mejer Nygårdsmoen ; Peter Arenfeldt Larsen ; Raghava Rao Mukkamala ; Benjamin Flesch ; Abid Hussain ; Ravi Vatrapu</i>	

TV RATINGS VS. SOCIAL MEDIA ENGAGEMENT: BIG SOCIAL DATA ANALYTICS OF THE SCANDINAVIAN TV TALK SHOW SKAVLAN	3849
<i>Henrikke Hovda Larsen ; Johanna Margareta Forsberg ; Sigrid Viken Hemstad ; Raghava Rao Mukkamala ; Abid Hussain ; Ravi Vatrapu</i>	
TOTALLY AUTOMATED KEYWORD EXTRACTION	3859
<i>Tayfun Pay</i>	
EFFICIENT NATURAL LANGUAGE PRE-PROCESSING FOR ANALYZING LARGE DATA SETS	3864
<i>Belaimine Billal ; Alessandro Fonseca ; Fatiha Sadat</i>	
A GRAPHEME-LEVEL APPROACH FOR CONSTRUCTING A KOREAN MORPHOLOGICAL ANALYZER WITHOUT LINGUISTIC KNOWLEDGE	3872
<i>Jihun Choi ; Jonghem Youn ; Sang-Goo Lee</i>	
LEXIDB: A SCALABLE CORPUS DATABASE MANAGEMENT SYSTEM	3880
<i>Matthew Coole ; Paul Rayson ; John Mariani</i>	
LARGE-SCALE TAXONOMY CATEGORIZATION FOR NOISY PRODUCT LISTINGS	3885
<i>Pradipto Das ; Yandi Xia ; Aaron Levine ; Giuseppe Di Fabbrizio ; Ankur Datta</i>	
SCALING CHARACTER-BASED MORPHOLOGICAL TAGGING TO FOURTEEN LANGUAGES	3895
<i>Georg Heigold ; Josef Van Genabith ; Günter Neumann</i>	
LIGHTWEIGHT SYSTEM FOR NE-TAGGED NEWS HEADLINES CORPUS CREATION	3903
<i>Avinash Kumar ; Dhaval Patel ; Nikita Jain</i>	
DOMAIN-SPECIFIC USER PREFERENCE PREDICTION BASED ON MULTIPLE USER ACTIVITIES	3913
<i>Yunfei Long ; Qin Lu ; Yue Xiao ; Minglei Li ; Chu-Ren Huang</i>	
DOCUMENT CLASSIFICATION THROUGH IMAGE-BASED CHARACTER EMBEDDING AND WILDCARD TRAINING	3922
<i>Daiki Shimada ; Ryunosuke Kotani ; Hitoshi Iyatomi</i>	
LARGE-SCALE TEXT PROCESSING PIPELINE WITH APACHE SPARK	3928
<i>A. Svyatkovskiy ; K. Imai ; M. Kroeger ; Y. Shiraito</i>	
AUTOMATIC CLASSIFICATION OF SECURITIES USING HIERARCHICAL CLUSTERING OF THE 10-KS	3936
<i>Hoseong Yang ; Hye Jin Lee ; Sungzoon Cho ; Eugene Cho</i>	
MAX-NODE SAMPLING: AN EXPANSION-DENSIFICATION ALGORITHM FOR DATA COLLECTION	3944
<i>Katchaguy Areekijseere ; Ricky Laishram ; Sucheta Soundarajan</i>	
REAL-TIME SENTIMENT ANALYSIS OF SAUDI DIALECT TWEETS USING SPARK	3947
<i>Adel Assiri ; Ahmed Emam ; Hmood Al-Dossari</i>	
MODELING, VALIDATION AND VERIFICATION OF CELL-SCAFFOLD CONTACT MEASUREMENTS OVER TERABYTE-SIZED 3D IMAGE COLLECTION	3951
<i>Peter Bajcsy ; Soweon Yoon ; Mylene Simon ; Mary Brady ; Ram Sriram ; Nathan Hotaling ; Nicholas Schaub ; Carl G. Simon ; Piotr M. Szczyplinski ; Stephen J. Florczyk</i>	
AN INTEGRATED ASSESSMENT APPROACH TO DIFFERENT COLLABORATIVE FILTERING ALGORITHMS	3954
<i>Raja Sarath Kumar Boddu</i>	
SEQUENTIAL RANDOMIZED MATRIX FACTORIZATION FOR GAUSSIAN PROCESSES	3957
<i>Shaunak D. Bopardikar ; George S. Eskander Ekladious</i>	
COMPARISON OF LOSSLESS VIDEO AND IMAGE COMPRESSION CODECS FOR MEDICAL COMPUTED TOMOGRAPHY DATASETS	3960
<i>Vy Bui ; Lin-Ching Chang ; Dunling Li ; Li-Yueh Hsu ; Marcus Y. Chen</i>	
ORANGE: SPATIAL BIG DATA ANALYSIS PLATFORM	3963
<i>Sunghwan Cho ; Sunghal Hong ; Changsoo Lee</i>	
ACCESSING AND DISTRIBUTING LARGE VOLUMES OF NETCDF DATA	3966
<i>Ranjeet Devarakonda ; Yaxing Wei ; Michele Thornton</i>	
NEXT-GEN TOOLS FOR BIG SCIENTIFIC DATA: ARM DATA CENTER EXAMPLE	3968
<i>Ranjeet Devarakonda ; Kyle Dumas ; Sheman Beus ; Everett Rush ; Bhargavi Krishna ; Rob Records ; Giri Prakash</i>	
CORRELATION BETWEEN WEATHER AND WEATHER-RELATED TWEETS — A PRELIMINARY STUDY	3971
<i>Srabasti Dutta ; Sumanro Ray ; S. Roy</i>	
FALL RECOGNITION USING WEARABLE TECHNOLOGIES AND MACHINE LEARNING ALGORITHMS	3974
<i>Austin Harris ; Hanna True ; Zhen Hu ; Jin Cho ; Nancy Fell ; Mina Sartipi</i>	

“WHAT MAKES A PRO EATING DISORDER HASHTAG”: USING HASHTAGS TO IDENTIFY PRO EATING DISORDER TUMBLR POSTS AND TWITTER USERS	3977
<i>Ling He ; Jiebo Luo</i>	
EVALUATION OF DISTRIBUTED PROCESSING OF CAFFE FRAMEWORK USING POOR PERFORMANCE DEVICE	3980
<i>Ayae Ichinose ; Masato Oguchi ; Atsuko Takefusa ; Hidemoto Nakada</i>	
FAST AND SPACE-EFFICIENT SECURE FREQUENT PATTERN MINING BY FHE.....	3983
<i>Hiroki Imabayashi ; Yu Ishimaki ; Akira Umayabara ; Hayato Yamana</i>	
ANALYSIS OF POKÉMON GO USING SOCIOPHYSICS APPROACH.....	3986
<i>Akira Ishii ; Masanori Ajito ; Yasuko Kawahata</i>	
PRIVACY-PRESERVING STRING SEARCH FOR GENOME SEQUENCES WITH FHE BOOTSTRAPPING OPTIMIZATION	3989
<i>Yu Ishimaki ; Hiroki Imabayashi ; Kana Shimizu ; Hayato Yamana</i>	
HARMONIZATION OF METHODS TO FACILITATE REPRODUCIBILITY IN MEDICAL DATA PROCESSING: APPLICATIONS TO DIFFUSION TENSOR MAGNETIC RESONANCE IMAGING	3992
<i>Jeffrey Jenkins ; Lin-Ching Chang ; Elizabeth Hutchinson ; M. Okan Irfanoglu ; Carlo Pierpaoli</i>	
TPR*-TREE PERFORMANCE IMPROVEMENT FOR BIG TACTICAL MOVING OBJECTS	3995
<i>Seungwoo Jeon ; Jaegi Hong ; Bonghee Hong ; Chumsu Kim</i>	
A DATA ANALYSIS AND VISUALIZATION SYSTEM FOR LARGE-SCALE E-BIKE DATA	3998
<i>Xiaoxia Jia ; Peng Cheng ; Jiming Chen</i>	
BIG DATA APPLICATION IN JOB TREND ANALYSIS	4001
<i>Priyanka Kale ; Shilpa Balan</i>	
NOWCASTING WITH SOCIAL MEDIA DATA	4004
<i>David L. Kimmey ; Jin S. Yoo</i>	
CAREERMAPPER: AN AUTOMATED RESUME EVALUATION TOOL	4005
<i>Vivian Lai ; Kyong Jin Shim ; Richard J. Oentaryo ; Philips K. Prasetyo ; Casey Vu ; Ee-Peng Lim ; David Lo</i>	
PREDICTED MAX DEGREE SAMPLING: SAMPLING IN DIRECTED NETWORKS TO MAXIMIZE NODE COVERAGE THROUGH CRAWLING	4008
<i>Ricky Laishram ; Katchaguy Areekijseree ; Sucheta Soundarajan</i>	
A GENERATOR OF TEST DATA SET FOR TACTICAL MOVING OBJECTS BASED ON VELOCITY.....	4011
<i>Jiwan Lee ; Jaegi Hong ; Bonghee Hong ; Jinsu Ahn</i>	
USING PARAPHRASES TO IMPROVE TWEET CLASSIFICATION: COMPARING WORDNET AND WORD EMBEDDING APPROACHES.....	4014
<i>Quanzhi Li ; Sameena Shah ; Mohammad Ghassemi ; Rui Fang ; Armineh Nourbakhsh ; Xiaomo Liu</i>	
A FRAMEWORK FOR LARGE-SCALE BACTERIAL MOTILITY BEHAVIOR ANALYSIS	4017
<i>Xiaomeng Liang ; Lin-Ching Chang ; Arash Massoudieh</i>	
INFERRING RELATIONS IN KNOWLEDGE GRAPHS WITH TENSOR DECOMPOSITIONS	4020
<i>Ankur Padia ; Konstantinos Kalpakis ; Tim Finin</i>	
TOWARDS A MORE METERLESS PARKING SYSTEM: UNDERSTANDING METER PAYMENT BEHAVIOR AND TRENDS IN WASHINGTON, DC.....	4023
<i>Benito O. Pérez ; Yiwei Ma ; Mengran Wang ; Xiaomeng Liang ; Negin Askarzadeh</i>	
HPC INFRASTRUCTURE TO SUPPORT THE NEXT-GENERATION ARM FACILITY DATA OPERATIONS.....	4026
<i>Giri Prakash ; Jitendra Kumar ; Everett Rush ; Robert Records ; Anthony Clodfelter ; Jimmy Voyles</i>	
USING AUTOMATED ENFORCEMENT DATA TO ACHIEVE VISION ZERO GOALS: A CASE STUDY	4029
<i>J. M. Rogers ; S. S. Dey ; R. Retting ; R. Jain ; X. Liang ; N. Askarzadeh</i>	
ANALYSIS OF TEAMWORK DIALOGUE: A DATA MINING APPROACH.....	4032
<i>Antonette Shibani ; Elizabeth Koh ; Vivian Lai ; Kyong Jin Shim</i>	
META-ANALYSIS OF BIG DATA SECURITY AND PRIVACY: SCHOLARLY LITERATURE GAPS	4035
<i>Kenneth David Strang ; Zhaohao Sun</i>	
AN APPROACH FOR EXTRACTING BIG MICRO-SCALE SEVERE WEATHER REGION TRAJECTORIES AUTOMATICALLY FROM METEOROLOGICAL RADAR DATA	4038
<i>Xingang Wang ; Zhigang Gai ; Suiping Qi</i>	
AN IMPROVED SOCIAL SPAMMER DETECTION BASED ON TRI-TRAINING	4040
<i>Guangxia Xu ; Jingteng Zhao ; Deling Huang</i>	
Author Index	