

2015 IEEE International Conference on Big Data (Big Data 2015)

**Santa Clara, California, USA
29 October – 1 November 2015**

Pages 1-726



**IEEE Catalog Number: CFP15BGD-POD
ISBN: 978-1-4799-9927-9**

**Copyright © 2015 by the Institute of Electrical and Electronic Engineers, Inc
All Rights Reserved**

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

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

******This publication is a representation of what appears in the IEEE Digital Libraries. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number: CFP15BGD-POD
ISBN (Print-On-Demand): 978-1-4799-9927-9

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

HOW BIG DATA CHANGES STATISTICAL MACHINE LEARNING	1
<i>Bottou, L.</i>	
MOVING PAST THE "WILD WEST" ERA FOR BIG DATA	2
<i>Jagadish, H.V.</i>	
CONQUERING BIG DATA WITH SPARK	3
<i>Stocia, I.</i>	
ONLINE AND ON-DEMAND PARTITIONING OF STREAMING GRAPHS	4
<i>Filippidou, I. ; Kotidis, Y.</i>	
LEARNING TO ACCURATELY COUNT WITH QUERY-DRIVEN PREDICTIVE ANALYTICS	14
<i>Anagnostopoulos, C. ; Triantafillou, P.</i>	
PRACTICAL MESSAGE-PASSING FRAMEWORK FOR LARGE-SCALE COMBINATORIAL OPTIMIZATION	24
<i>Inho Cho ; Soya Park ; Sejun Park ; Dongsu Han ; Jinwoo Shin</i>	
REWRITING COMPLEX SPARQL ANALYTICAL QUERIES FOR EFFICIENT CLOUD-BASED PROCESSING	32
<i>Ravindra, P. ; HyeongSik Kim ; Anyanwu, K.</i>	
CONCEPT HIERARCHIES AND HUMAN NAVIGATION	38
<i>Aguinaga, S. ; Nambiar, A. ; Zuozhu Liu ; Weninger, T.</i>	
ITERATIVELY REFINING SVMS USING PRIORS	46
<i>de Fortuny, E.J. ; Evgeniou, T. ; Martens, D. ; Provost, F.</i>	
TOWARDS SCALABLE QUANTILE REGRESSION TREES	53
<i>Bhat, Harish S. ; Kumar, Nitesh ; Vaz, Garnet J.</i>	
SUPER-CWC AND SUPER-LCC: SUPER FAST FEATURE SELECTION ALGORITHMS	61
<i>Shin, K. ; Kuboyama, T. ; Hashimoto, T. ; Shepard, D.</i>	
CONSIDERATIONS AND RECOMMENDATIONS FOR DATA AVAILABILITY FOR DATA ANALYTICS FOR MANUFACTURING	68
<i>Libes, D. ; Seungjun Shin ; Jungyub Woo</i>	
SCALEGRAPH: A HIGH-PERFORMANCE LIBRARY FOR BILLION-SCALE GRAPH ANALYTICS	76
<i>Suzumura, T. ; Ueno, K.</i>	
SYSTEM AND ARCHITECTURE LEVEL CHARACTERIZATION OF BIG DATA APPLICATIONS ON BIG AND LITTLE CORE SERVER ARCHITECTURES	85
<i>Malik, M. ; Rafatirah, S. ; Sasan, A. ; Homayoun, H.</i>	
DATA STREAMING ALGORITHMS FOR THE KOLMOGOROV-SMIRNOV TEST	95
<i>Lall, A.</i>	
TECHNIQUES FOR FAST AND SCALABLE TIME SERIES TRAFFIC GENERATION	105
<i>Jilong Kuang ; Waddington, D.G. ; Changhui Lin</i>	
ENERGY-EFFICIENT ACCELERATION OF BIG DATA ANALYTICS APPLICATIONS USING FPGAS	115
<i>Neshatpour, K. ; Malik, M. ; Ghodrat, M.A. ; Sasan, A. ; Homayoun, H.</i>	
WORKLOAD SCHEDULING IN DISTRIBUTED STREAM PROCESSORS USING GRAPH PARTITIONING	124
<i>Fischer, L. ; Bernstein, A.</i>	
EVALUATING DIFFERENT DISTRIBUTED-CYBER-INFRASTRUCTURE FOR DATA AND COMPUTE INTENSIVE SCIENTIFIC APPLICATION	134
<i>Das, Arghya Kusum ; Park, Seung-Jong ; Hong, Jaeki ; Chang, Wooseok</i>	
SCALEJOIN: A DETERMINISTIC, DISJOINT-PARALLEL AND SKEW-RESILIENT STREAM JOIN	144
<i>Gulisano, V. ; Nikolakopoulos, Y. ; Papatriantafilou, M. ; Tsigas, P.</i>	
WHEN COMPUTING MEETS HETEROGENEOUS CLUSTER: WORKLOAD ASSIGNMENT IN GRAPH COMPUTATION	154
<i>Jilong Xue ; Zhi Yang ; Shian Hou ; Yafei Dai</i>	
A SCALABLE PARALLEL XQUERY PROCESSOR	164
<i>Preston Carman, E. ; Westmann, T. ; Borkar, V.R. ; Carey, M.J. ; Tsotras, V.J.</i>	
COMPUTING LOAD AWARE AND LONG-VIEW LOAD BALANCING FOR CLUSTER STORAGE SYSTEMS	174
<i>Guoxin Liu ; Haiying Shen ; Haoyu Wang</i>	

DISTRIBUTED FRANK-WOLFE UNDER PIPELINED STALE SYNCHRONOUS PARALLELISM.....	184
<i>Tran, Nam-Luc ; Peel, Thomas ; Skhiri, Sabri</i>	
EVALUATING CLOUD FRAMEWORKS ON GENOMIC APPLICATIONS.....	193
<i>Bertoni, M. ; Ceri, S. ; Kaitoua, A. ; Pinoli, P.</i>	
TOWARDS GREEN CLOUD COMPUTING: DEMAND ALLOCATION AND PRICING POLICIES FOR CLOUD SERVICE BROKERAGE	203
<i>Chenxi Qiu ; Haiying Shen ; Lihua Chen</i>	
ELASTIC COMPLEX EVENT PROCESSING EXPLOITING PREDICTION.....	213
<i>Zacheilas, N. ; Kalogeraki, V. ; Zygouras, N. ; Panagiotou, N. ; Gunopulos, D.</i>	
PORTHADOOP: SUPPORT DIRECT HPC DATA PROCESSING IN HADOOP.....	223
<i>Xi Yang ; Ning Liu ; Bo Feng ; Xian-He Sun ; Shujia Zhou</i>	
MACHINE LEARNING AT THE LIMIT	233
<i>Canny, J. ; Huasha Zhao ; Jaros, B. ; Ye Chen ; Jiangchang Mao</i>	
PERFORMANCE CHARACTERIZATION AND ACCELERATION OF IN-MEMORY FILE SYSTEMS FOR HADOOP AND SPARK APPLICATIONS ON HPC CLUSTERS	243
<i>Islam, N.S. ; Wasi-ur-Rahman, M. ; Xiaoyi Lu ; Shankar, D. ; Panda, D.K.</i>	
PANOPTICON: A LOCK BROKER ARCHITECTURE FOR SCALABLE TRANSACTIONS IN THE DATACENTER.....	253
<i>Tasci, S. ; Demirbas, M.</i>	
TOWARD LOCALITY-AWARE SCHEDULING FOR CONTAINERIZED CLOUD SERVICES.....	263
<i>Dongfang Zhao ; Mandagere, N. ; Alatorre, G. ; Mohamed, M. ; Ludwig, H.</i>	
ATOM: AUTOMATED TRACKING, ORCHESTRATION AND MONITORING OF RESOURCE USAGE IN INFRASTRUCTURE AS A SERVICE SYSTEMS	271
<i>Min Du ; Feifei Li</i>	
COMPOSABLE AND EFFICIENT FUNCTIONAL BIG DATA PROCESSING FRAMEWORK.....	279
<i>Dongyao Wu ; Sakr, S. ; Liming Zhu ; Qinghua Lu</i>	
HYBRID ACTIVE LEARNING FOR NON-STATIONARY STREAMING DATA WITH ASYNCHRONOUS LABELING	287
<i>Hyunjoo Kim ; Madhvanath, S. ; Tong Sun</i>	
OCTOPUS: A MULTI-JOB SCHEDULER FOR GRAPHLAB	293
<i>Padala, S. ; Kumar, D. ; Raj, A. ; Dharanipragada, J.</i>	
SPARK DEPLOYMENT AND PERFORMANCE EVALUATION ON THE MARENOSTRUM SUPERCOMPUTER.....	299
<i>Tous, Ruben ; Gounaris, Anastasios ; Tripiana, Carlos ; Torres, Jordi ; Girona, Sergi ; Ayguade, Eduard ; Labarta, Jesus ; Becerra, Yolanda ; Carrera, David ; Valero, Mateo</i>	
G-STORM: GPU-ENABLED HIGH-THROUGHPUT ONLINE DATA PROCESSING IN STORM	307
<i>Zhenhua Chen ; Jielong Xu ; Jian Tang ; Kwiat, K. ; Kamhoua, C.</i>	
CHRONOS: FAILURE-AWARE SCHEDULING IN SHARED HADOOP CLUSTERS.....	313
<i>Yildiz, O. ; Ibrahim, S. ; Phuong, T.A. ; Antoniu, G.</i>	
AN ARCHITECTURE FOR STREAM OLAP EXPLOITING SPE AND OLAP ENGINE	319
<i>Nakabasami, K. ; Amagasa, T. ; Shaikh, S.A. ; Gass, F. ; Kitagawa, H.</i>	
TWO-MODE DATA DISTRIBUTION SCHEME FOR HETEROGENEOUS STORAGE IN DATA CENTERS	327
<i>Wei Xie ; Jiang Zhou ; Reyes, M. ; Noble, J. ; Yong Chen</i>	
A PREDICTIVE SCHEDULING FRAMEWORK FOR FAST AND DISTRIBUTED STREAM DATA PROCESSING	333
<i>Teng Li ; Jian Tang ; Jielong Xu</i>	
A SCALABLE IMPLEMENTATION OF INFORMATION THEORETIC FEATURE SELECTION FOR HIGH DIMENSIONAL DATA	339
<i>Kleerekoper, A. ; Pappas, M. ; Pocock, A. ; Brown, G. ; Lujan, M.</i>	
EDGE IMPORTANCE IDENTIFICATION FOR ENERGY EFFICIENT GRAPH PROCESSING	347
<i>Faisal, S.M. ; Tziantzioulis, G. ; Gok, A.M. ; Hardavellas, N. ; Ogrenci-Memik, S. ; Parthasarathy, S.</i>	
REGULAR EXPRESSION ACCELERATION ON THE MICRON AUTOMATA PROCESSOR: BRILL TAGGING AS A CASE STUDY	355
<i>Zhou, K. ; Wadden, J. ; Fox, J.J. ; Ke Wang ; Brown, D.E. ; Skadron, K.</i>	
PARALLEL IN-MEMORY TRAJECTORY-BASED SPATIOTEMPORAL TOPOLOGICAL JOIN	361
<i>Ray, S. ; Demke Brown, A. ; Koudas, N. ; Blanco, R. ; Goel, A.K.</i>	
SPATIALLY CLUSTERED JOIN ON HETEROGENEOUS SCIENTIFIC DATA SETS.....	371
<i>Bin Dong ; Byna, S. ; Kesheng Wu</i>	
RECOMMENDING MISSING SENSOR VALUES	381
<i>Chung-Yi Li ; Wei-Lun Su ; McKenzie, T.G. ; Fu-Chun Hsu ; Shou-De Lin ; Hsu, J.Y.-J. ; Gibbons, P.B.</i>	

THE ROLES OF NETWORK COMMUNITIES IN SOCIAL INFORMATION DIFFUSION	391
<i>Cheng-Te Li ; Yu-Jen Lin ; Mi-Yen Yeh</i>	
BIG DATA ENTITY RESOLUTION: FROM HIGHLY TO SOMEHOW SIMILAR ENTITY DESCRIPTIONS IN THE WEB	401
<i>Efthymiou, V. ; Stefanidis, K. ; Christophides, V.</i>	
PARALLEL META-BLOCKING: REALIZING SCALABLE ENTITY RESOLUTION OVER LARGE, HETEROGENEOUS DATA	411
<i>Efthymiou, V. ; Papadakis, G. ; Papastefanatos, G. ; Stefanidis, K. ; Palpanas, T.</i>	
SLINGSHOT: A MODULAR FRAMEWORK FOR DESIGNING DATA PROCESSING SYSTEMS	421
<i>Simion, B. ; Ilha, D.N. ; Ray, S. ; Barron, L. ; Demke Brown, A. ; Johnson, R.</i>	
LABBOOK: METADATA-DRIVEN SOCIAL COLLABORATIVE DATA ANALYSIS	431
<i>Kandogan, E. ; Roth, M. ; Schwarz, P. ; Hui, J. ; Terrizzano, I. ; Christodoulakis, C. ; Miller, R.J.</i>	
TRUSTMR: COMPUTATION INTEGRITY ASSURANCE SYSTEM FOR MAPREDUCE	441
<i>Ulusoy, H. ; Kantarcioglu, M. ; Pattuk, E.</i>	
ACCOUNTABLEMR: TOWARD ACCOUNTABLE MAPREDUCE SYSTEMS	451
<i>Ulusoy, H. ; Kantarcioglu, M. ; Pattuk, E. ; Kagal, L.</i>	
TKSIMGPU: A PARALLEL TOP-K TRAJECTORY SIMILARITY QUERY PROCESSING ALGORITHM FOR GPGPUS	461
<i>Leal, E. ; Gruenwald, L. ; Jianting Zhang ; You, S.</i>	
A TRANSACTION MODEL FOR MANAGEMENT OF REPLICATED DATA WITH MULTIPLE CONSISTENCY LEVELS	470
<i>Tripathi, A. ; Thirunavukarasu, B.D.</i>	
QUADTREE-BASED LIGHTWEIGHT DATA COMPRESSION FOR LARGE-SCALE GEOSPATIAL RASTERS ON MULTI-CORE CPUS	478
<i>Jianting Zhang ; Simin You ; Le Gruenwald</i>	
DSDQUERY DSI - QUERYING SCIENTIFIC DATA REPOSITORIES WITH STRUCTURED OPERATORS	485
<i>Ebenstein, R. ; Agrawal, G.</i>	
BROWN DOG: LEVERAGING EVERYTHING TOWARDS AUTOCURATION	493
<i>Padhy, S. ; Jansen, G. ; Alameda, J. ; Black, E. ; Diesendruck, L. ; Dietze, M. ; Kumar, P. ; Kooper, R. ; Jong Lee ; Rui Liu ; Marciano, R. ; Marini, L. ; Mattson, D. ; Minsker, B. ; Navarro, C. ; Slavenas, M. ; Sullivan, W. ; Votava, J. ; Zharnitsky, I. ; McHenry, K.</i>	
COST-EFFICIENT PARTITIONING OF SPATIAL DATA ON CLOUD	501
<i>Akdogan, A. ; Indrakanti, S. ; Demiryurek, U. ; Shahabi, C.</i>	
BIGFUN: A PERFORMANCE STUDY OF BIG DATA MANAGEMENT SYSTEM FUNCTIONALITY	507
<i>Pirzadeh, P. ; Carey, M.J. ; Westmann, T.</i>	
A FLEXIBLE QOS FORTIFIED DISTRIBUTED KEY-VALUE STORAGE SYSTEM FOR THE CLOUD	515
<i>Tonglin Li ; Ke Wang ; Dongfang Zhao ; Kan Qiao ; Sadooghi, I. ; Xiaobing Zhou ; Raicu, I.</i>	
TPS: A TASK PLACEMENT STRATEGY FOR BIG DATA WORKFLOWS	523
<i>Ebrahimi, M. ; Mohan, A. ; Shiyong Lu ; Reynolds, R.</i>	
IMPROVING TRANSACTION PROCESSING PERFORMANCE BY CONSENSUS REDUCTION	531
<i>Yuqing Zhu ; Yilei Wang</i>	
BENCHMARKING KEY-VALUE STORES ON HIGH-PERFORMANCE STORAGE AND INTERCONNECTS FOR WEB-SCALE WORKLOADS	539
<i>Shankar, D. ; Xiaoyi Lu ; Wasi-ur-Rahman, M. ; Islam, N. ; Panda, D.K.</i>	
AN ITERATIVE METHODOLOGY FOR BIG DATA MANAGEMENT, ANALYSIS AND VISUALIZATION	545
<i>Tardio, R. ; Mate, A. ; Trujillo, J.</i>	
BANDWIDTH-EFFICIENT DISTRIBUTED K-NEAREST-NEIGHBOR SEARCH WITH DYNAMIC TIME WARPING	551
<i>Chin-Chi Hsu ; Perng-Hwa Kung ; Mi-Yen Yeh ; Shou-De Lin ; Gibbons, P.B.</i>	
DYNAMIC THEME TRACKING IN TWITTER	561
<i>Liang Zhao ; Feng Chen ; Chang-Tien Lu ; Ramakrishnan, N.</i>	
SYNTACTICDIFF: OPERATOR-BASED TRANSFORMATION FOR COMPARATIVE TEXT MINING	571
<i>Massung, S. ; ChengXiang Zhai</i>	
VISUAL ANALYSIS OF BI-DIRECTIONAL MOVEMENT BEHAVIOR	581
<i>Yixian Zheng ; Wenchao Wu ; Huamin Qu ; Chunyan Ma ; Ni, L.M.</i>	
USER-CURATED IMAGE COLLECTIONS: MODELING AND RECOMMENDATION	591
<i>Yuncheng Li ; Yang Cong ; Tao Mei ; Jiebo Luo</i>	

ANGULAR QUANTIZATION BASED AFFINITY PROPAGATION CLUSTERING AND ITS APPLICATION TO ASTRONOMICAL BIG SPECTRA DATA	601
<i>Ke Wang ; Ping Guo ; A-Li Luo</i>	
SCALABLE CLASSIFICATION FOR LARGE DYNAMIC NETWORKS	609
<i>Yibo Yao ; Holder, L.B.</i>	
CINTIA: A DISTRIBUTED, LOW-LATENCY INDEX FOR BIG INTERVAL DATA	619
<i>Mavlyutov, R. ; Cudre-Mauroux, P.</i>	
REVEALING THE FOG-OF-WAR: A VISUALIZATION-DIRECTED, UNCERTAINTY-AWARE APPROACH FOR EXPLORING HIGH-DIMENSIONAL DATA	629
<i>Yang Wang ; Kwan-Liu Ma</i>	
INFERRING CROWD-SOURCED VENUES FOR TWEETS	639
<i>Bokai Cao ; Chen, F. ; Joshi, D. ; Yu, P.S.</i>	
CORE DECOMPOSITION IN LARGE TEMPORAL GRAPHS	649
<i>Huanhuan Wu ; Cheng, J. ; Yi Lu ; Yiping Ke ; Yuzhen Huang ; Da Yan ; Hejun Wu</i>	
RECOMMENDING FORUM POSTS TO DESIGNATED EXPERTS	659
<i>Cho, J.H.D. ; Yanen Li ; Girju, R. ; ChengXiang Zhai</i>	
ACCELERATING COLLABORATIVE FILTERING USING CONCEPTS FROM HIGH PERFORMANCE COMPUTING	667
<i>Gates, M. ; Anzt, H. ; Kurzak, J. ; Dongarra, J.</i>	
MODELLING CASCADES OVER TIME IN MICROBLOGS	677
<i>Wei Xie ; Zhu, F. ; Siyuan Liu ; Ke Wang</i>	
CSFINDER: A COLD-START FRIEND FINDER IN LARGE-SCALE SOCIAL NETWORKS	687
<i>Salem, Y. ; Jun Hong ; Liu, W.</i>	
EFFECTIVELY CROWDSOURCING THE ACQUISITION AND ANALYSIS OF VISUAL DATA FOR DISASTER RESPONSE	697
<i>Hien To ; Seon Ho Kim ; Shahabi, C.</i>	
FULL DIFFUSION HISTORY RECONSTRUCTION IN NETWORKS	707
<i>Zhen Chen ; Hanghang Tong ; Lei Ying</i>	
ADAM: AN ADAPTIVE MONITORING FRAMEWORK FOR SAMPLING AND FILTERING ON IOT DEVICES	717
<i>Trihinas, D. ; Pallis, G. ; Dikaiaikos, M.D.</i>	
MODELING GRAPHS USING A MIXTURE OF KRONECKER MODELS	727
<i>Mahapatra, S. ; Chandola, V.</i>	
DATA QUALITY ASSESSMENT AND ANOMALY DETECTION VIA MAP/REDUCE AND LINKED DATA: A CASE STUDY IN THE MEDICAL DOMAIN	737
<i>Bonner, S. ; McGough, A.S. ; Kureshi, I. ; Brennan, J. ; Theodoropoulos, G. ; Moss, L. ; Corsar, D. ; Antoniou, G.</i>	
SIGCO: MINING SIGNIFICANT CORRELATIONS VIA A DISTRIBUTED REAL-TIME COMPUTATION ENGINE	747
<i>Tian Guo ; Calbimonte, J.-P. ; Hao Zhuang ; Aberer, K.</i>	
IDENTIFYING SMALLEST UNIQUE SUBGRAPHS IN A HETEROGENEOUS SOCIAL NETWORK	757
<i>Wang, Yen-Kai ; Chen, Wei-Ming ; Li, Cheng-Te ; Lin, Shou-De</i>	
TOWARD PRECISE USER-TOPIC ALIGNMENT IN ONLINE SOCIAL MEDIA	767
<i>Jiejun Xu ; Tsai-Ching Lu</i>	
VISUAL INTERFACE FOR EXPLORING CAUTION SPOTS FROM VEHICLE RECORDER BIG DATA	776
<i>Itoh, M. ; Yokoyama, D. ; Toyoda, M. ; Kitsuregawa, M.</i>	
ACURDION: AN ADAPTIVE CLUSTERING-BASED ALGORITHM FOR TRACING LARGE-SCALE MPI APPLICATIONS	785
<i>Bahmani, A. ; Mueller, F.</i>	
TIME MAPS: A TOOL FOR VISUALIZING MANY DISCRETE EVENTS ACROSS MULTIPLE TIMESCALES	793
<i>Watson, M.C.</i>	
LEARNING RELEVANCE FROM CLICK DATA VIA NEURAL NETWORK BASED SIMILARITY MODELS	801
<i>Xugang Ye ; Zijie Qi ; Massey, D.</i>	
MATISSE: A VISUAL ANALYTICS SYSTEM FOR EXPLORING EMOTION TRENDS IN SOCIAL MEDIA TEXT STREAMS	807
<i>Steed, C.A. ; Drouhard, M. ; Beaver, J. ; Pyle, J. ; Bogen, P.L.</i>	
ROBUST CROWD BIAS CORRECTION VIA DUAL KNOWLEDGE TRANSFER FROM MULTIPLE OVERLAPPING SOURCES	815
<i>Sihong Xie ; Qingbo Hu ; Jingyuan Zhang ; Jing Gao ; Wei Fan ; Yu, P.S.</i>	

A COMMUNITY DRIVEN SOCIAL RECOMMENDATION SYSTEM	821
<i>Labwani, D. ; Somayajulu, D.V.L.N. ; Krishna, P.R.</i>	
TASK-BASED RECOMMENDATION ON A WEB-SCALE	827
<i>Yongfeng Zhang ; Min Zhang ; Yiqun Liu ; Chua Tat-Seng ; Yi Zhang ; Shaoping Ma</i>	
MULTI-MODAL LEARNING FOR VIDEO RECOMMENDATION BASED ON MOBILE APPLICATION USAGE	837
<i>Xiaowei Jia ; Aosen Wang ; Xiaoyi Li ; Guangxu Xun ; Wenyao Xu ; Aidong Zhang</i>	
IMPROVING EEG FEATURE LEARNING VIA SYNCHRONIZED FACIAL VIDEO	843
<i>Xiaoyi Li ; Xiaowei Jia ; Guangxu Xun ; Aidong Zhang</i>	
MMC-MARGIN: IDENTIFICATION OF MAXIMUM FREQUENT SUBGRAPHS BY METROPOLIS MONTE CARLO SAMPLING	849
<i>Muyi Liu ; Gribskov, M.</i>	
KEYLABEL ALGORITHMS FOR KEYWORD SEARCH IN LARGE GRAPHS	857
<i>Yue Wang ; Ke Wang ; Fu, A.W.-C. ; Wong, R.C.-W.</i>	
SPATIO-TEMPORAL ASYNCHRONOUS CO-OCCURRENCE PATTERN FOR BIG CLIMATE DATA TOWARDS LONG-LEAD FLOOD PREDICTION	865
<i>Chung-Hsien Yu ; Dong Luo ; Wei Ding ; Cohen, J. ; Small, D. ; Islam, S.</i>	
USING BIG DATA TO STUDY THE LINK BETWEEN HUMAN MOBILITY AND SOCIO-ECONOMIC DEVELOPMENT	871
<i>Pappalardo, L. ; Smoreda, Z. ; Pedreschi, D. ; Giannotti, F.</i>	
CLUSTER-BASED AGGREGATE FORECASTING FOR RESIDENTIAL ELECTRICITY DEMAND USING SMART METER DATA	879
<i>Wijaya, T.K. ; Vasirani, M. ; Humeau, S. ; Aberer, K.</i>	
A SCALABLE APPROACH FOR DATA-DRIVEN TAXI RIDE-SHARING SIMULATION	888
<i>Ota, Masayo ; Vo, Huy ; Silva, Claudio ; Freire, Juliana</i>	
EVERYONECOUNTS: DATA-DRIVEN DIGITAL ADVERTISING WITH UNCERTAIN DEMAND MODEL IN METRO NETWORKS	898
<i>Desheng Zhang ; Ritobing Jiang ; Shii'ai Wang ; Yanmin Zhu ; Bo Yang ; Jian Cao ; Fan Zhang ; Tian He</i>	
FAST DECENTRALIZED GRADIENT DESCENT METHOD AND APPLICATIONS TO IN-SITU SEISMIC TOMOGRAPHY	908
<i>Liang Zhao ; Wen-Zhan Song ; Xiaojing Ye</i>	
SCIENTIFIC COMPUTING MEETS BIG DATA TECHNOLOGY: AN ASTRONOMY USE CASE	918
<i>Zhao Zhang ; Barbary, K. ; Nothaft, F.A. ; Sparks, E. ; Zahn, O. ; Franklin, M.J. ; Patterson, D.A. ; Perlmutter, S.</i>	
AN INTERACTIVE LEARNING FRAMEWORK FOR SCALABLE CLASSIFICATION OF PATHOLOGY IMAGES	928
<i>Nalitsnik, Michael ; Gutman, David A ; Kong, Jun ; Cooper, Lee A D</i>	
AMERICA TWEETS CHINA: A FINE-GRAINED ANALYSIS OF THE STATE AND INDIVIDUAL CHARACTERISTICS REGARDING ATTITUDES TOWARDS CHINA	936
<i>Yu Wang ; Jianbo Yuan ; Jiebo Luo</i>	
A DATA-DRIVEN APPROACH TO EXTRACT CONNECTIVITY STRUCTURES FROM DIFFUSION TENSOR IMAGING DATA	944
<i>Yu Jin ; JaJa, J.F. ; Rong Chen ; Herskovits, E.H.</i>	
A MAPREDUCE BASED K-NN JOINS PROBABILISTIC CLASSIFIER	952
<i>Chatzigeorgakidis, G. ; Karagiorgou, S. ; Athanasiou, S. ; Skiadopoulos, S.</i>	
SCALABLE K-NN BASED TEXT CLUSTERING	958
<i>Lulli, A. ; Debatty, T. ; Dell'Amico, M. ; Michiardi, P. ; Ricci, L.</i>	
AN ENSEMBLE LEARNING BASED APPROACH FOR BUILDING AIRFARE FORECAST SERVICE	964
<i>Yuwen Chen ; Jian Cao ; Shanshan Feng ; Yudong Tan</i>	
NEXT-TERM STUDENT GRADE PREDICTION	970
<i>Sweeney, M. ; Lester, J. ; Rangwala, H.</i>	
PREDICTING THE LOCATION OF USERS ON TWITTER FROM LOW DENSITY GRAPHS	976
<i>Apreleva, S. ; Cantarero, A.</i>	
HOW NOT TO DROWN IN A SEA OF INFORMATION: AN EVENT RECOGNITION APPROACH	984
<i>Alevizos, E. ; Artikis, A. ; Patroumpas, K. ; Vodas, M. ; Theodoridis, Y. ; Pelekis, N.</i>	
SMOG DISASTER FORECASTING USING SOCIAL WEB DATA AND PHYSICAL SENSOR DATA	991
<i>Jiaoyan Chen ; Huajun Chen ; Daning Hu ; Pan, J.Z. ; Yalin Zhou</i>	
LARGE SCALE SUPPORT VECTOR REGRESSION FOR AVIATION SAFETY	999
<i>Das, K. ; Bhaduri, K. ; Matthews, B.L. ; Oza, N.C.</i>	

CITY USERS' CLASSIFICATION WITH MOBILE PHONE DATA	1007
<i>Gabrielli, L. ; Furletti, B. ; Trasarti, R. ; Giannotti, F. ; Pedreschi, D.</i>	
SPALER: SPARK AND GRAPHX BASED DE NOVO GENOME ASSEMBLER	1013
<i>Abu-Doleh, A. ; Catalyurek, U.V.</i>	
TRAFFIC FORECASTING IN COMPLEX URBAN NETWORKS: LEVERAGING BIG DATA AND MACHINE LEARNING	1019
<i>Schimbinschi, F. ; Xuan Vinh Nguyen ; Bailey, J. ; Leckie, C. ; Hai Vu ; Kotagiri, R.</i>	
PREDICTION OF PHYSIOLOGICAL SUBSYSTEM FAILURE AND ITS IMPACT IN THE PREDICTION OF PATIENT MORTALITY	1025
<i>Barajas, K.C. ; Akella, R.</i>	
EFFICIENT DISTRIBUTED MAXIMUM MATCHING FOR SOLVING THE CONTAINER EXCHANGE PROBLEM IN THE MARITIME INDUSTRY	1031
<i>Fei Shao ; Li-Yung Ho ; Jan-Jan Wu ; Pangfeng Liu</i>	
CELL ANALYTICS IN COMPOUND HIT SELECTION OF BACTERIAL INHIBITORS	1037
<i>Trevino, R.P. ; Kawamoto, S.A. ; Lamkin, T.J. ; Huan Liu</i>	
MINING TARGET USERS FOR ONLINE MARKETING BASED ON APP STORE DATA	1043
<i>Xiuqiang He ; Wenyuan Dai ; Guoxiang Cao ; Ruiming Tang ; Mingxuan Yuan ; Qiang Yang</i>	
SCALABLE COMMUNITY DISCOVERY FROM MULTI-FACETED GRAPHS	1053
<i>Metwally, A. ; Jia-Yu Pan ; Minh Doan ; Faloutsos, C.</i>	
TOWARDS REAL-TIME CUSTOMER EXPERIENCE PREDICTION FOR TELECOMMUNICATION OPERATORS	1063
<i>Diaz-Aviles, E. ; Pinelli, F. ; Lynch, K. ; Nabi, Z. ; Gkoufas, Y. ; Bouillet, E. ; Calabrese, F. ; Coughlan, E. ; Holland, P. ; Salzwedel, J.</i>	
EARLY EXPERIENCE WITH OPTIMIZING I/O PERFORMANCE USING HIGH-PERFORMANCE SSDS FOR IN-MEMORY CLUSTER COMPUTING	1073
<i>Choi, I.S. ; Weiqing Yang ; Yang-Suk Kee</i>	
AN EVALUATION OF ALTERNATIVE SHARED-NOTHING ARCHITECTURE FOR ANALYTICAL PROCESSING SYSTEMS	1084
<i>Hyunsik Choi ; Jongyoung Park ; Yong In Lee ; Kangho Roh ; Kwanghyun La</i>	
CONTROLLED EXPERIMENTS FOR DECISION-MAKING IN E-COMMERCE SEARCH	1094
<i>Goswami, A. ; Wei Han ; Zhenrui Wang ; Jiang, A.</i>	
SEMANTICS FOR BIG DATA ACCESS & INTEGRATION: IMPROVING INDUSTRIAL EQUIPMENT DESIGN THROUGH INCREASED DATA USABILITY	1103
<i>Williams, J.W. ; Cuddihy, P. ; McHugh, J. ; Aggour, K.S. ; Menon, A. ; Gustafson, S.M. ; Healy, T.</i>	
ONLINE ANOMALY DETECTION OVER BIG DATA STREAMS	1113
<i>Rettig, L. ; Khayati, M. ; Cudre-Mauroux, P. ; Piorkowski, M.</i>	
CONTEXTUAL VERIFICATION FOR FALSE ALARM REDUCTION IN MARITIME ANOMALY DETECTION	1123
<i>Radon, A.N. ; Ke Wang ; Glasser, U. ; Wehn, H. ; Westwell-Roper, A.</i>	
BATCH-MODE ACTIVE LEARNING FOR TECHNOLOGY-ASSISTED REVIEW	1134
<i>Saha, T.K. ; Al Hasan, M. ; Burgess, C. ; Ahsan Habib, M. ; Johnson, J.</i>	
A PIPELINE FOR EXTRACTING AND DEDUPLICATING DOMAIN-SPECIFIC KNOWLEDGE BASES	1144
<i>Kejriwal, M. ; Qiaoling Liu ; Jacob, F. ; Javed, F.</i>	
EXOS: EXPANSION ON SESSION FOR ENHANCING EFFECTIVENESS OF QUERY AUTO-COMPLETION	1154
<i>Fang-Hsiang Su ; Somaiya, M. ; Mishra, S. ; Mukherjee, R.</i>	
PROBABILISTIC K^M-ANONYMITY EFFICIENT ANONYMIZATION OF LARGE SET-VALUED DATASETS	1164
<i>Acs, G. ; Achara, J.P. ; Castelluccia, C.</i>	
ADMM BASED SCALABLE MACHINE LEARNING ON SPARK	1174
<i>Dhar, S. ; Congrui Yi ; Ramakrishnan, N. ; Shah, M.</i>	
RECORD-AWARE COMPRESSION FOR BIG TEXTUAL DATA ANALYSIS ACCELERATION	1183
<i>Dapeng Dong ; Herbert, J.</i>	
GRAPH ANALYTICS USING VERTICAL RELATIONAL DATABASE	1191
<i>Jindal, A. ; Madden, S. ; Castellanos, M. ; Meichum Hsu</i>	
AUTOMOTIVE BIG DATA: APPLICATIONS, WORKLOADS AND INFRASTRUCTURES	1201
<i>Luckow, A. ; Kennedy, K. ; Manhardt, F. ; Djerekarov, E. ; Vorster, B. ; Apon, A.</i>	
COST-SENSITIVE OPTIMIZATION OF AUTOMATED INSPECTION	1211
<i>Cinar, G.T. ; Thompson, J. ; Srinivasan, S.</i>	

FROM PERFORMANCE PROFILING TO PREDICTIVE ANALYTICS WHILE EVALUATING HADOOP COST-EFFICIENCY IN ALOJA	1220
<i>Poggi, N. ; Berral, J.L. ; Carrera, D. ; Call, A. ; Gagliardi, F. ; Reinauer, R. ; Vujic, N. ; Green, D. ; Blakeley, J.</i>	
QUERY SENSE DISAMBIGUATION LEVERAGING LARGE SCALE USER BEHAVIORAL DATA.....	1230
<i>Korayem, M. ; Ortiz, C. ; AlJadda, K. ; Grainger, T.</i>	
PERSONALIZED EXPERTISE SEARCH AT LINKEDIN	1238
<i>Viet Ha-Thuc ; Venkataraman, G. ; Rodriguez, M. ; Sinha, S. ; Sundaram, S. ; Lin Guo</i>	
HOW VALUABLE IS YOUR DATA? A QUANTITATIVE APPROACH USING DATA MINING	1248
<i>Deolalikar, V.</i>	
MINING LIFESTYLE PERSONAS AT SCALE IN E-COMMERCE.....	1254
<i>Kang Li ; Deolalikar, V. ; Pradhan, N.</i>	
SDFS: SECURE DISTRIBUTED FILE SYSTEM FOR DATA-AT-REST SECURITY FOR HADOOP-AS-A-SERVICE	1262
<i>Zerfos, P. ; Hangu Yeo ; Paulovicks, B.D. ; Sheinin, V.</i>	
OPEN RESEARCH CHALLENGES WITH BIG DATA - A DATA-SCIENTIST'S PERSPECTIVE	1272
<i>Sukumar, S.R.</i>	
MARITIME SITUATION ANALYSIS FRAMEWORK: VESSEL INTERACTION CLASSIFICATION AND ANOMALY DETECTION.....	1279
<i>Shahir, H.Y. ; Glasser, U. ; Shahir, A.Y. ; Wehn, H.</i>	
PAIRS: A SCALABLE GEO-SPATIAL DATA ANALYTICS PLATFORM	1290
<i>Klein, L.J. ; Marianno, F.J. ; Albrecht, C.M. ; Freitag, M. ; Siyuan Lu ; Hinds, N. ; Xiaoyan Shao ; Bermudez Rodriguez, S. ; Hamann, H.F.</i>	
POST-PURCHASE RECOMMENDATIONS IN LARGE-SCALE ONLINE MARKETPLACES	1299
<i>Katukuri, J. ; Konik, T. ; Mukherjee, R. ; Kolay, S.</i>	
REVENUE MAXIMIZATION FOR TELECOMMUNICATIONS COMPANY WITH SOCIAL VIRAL MARKETING.....	1306
<i>Hong-Han Shuai ; Chih-Ya Shen ; Hsiang-Chun Hsu ; De-Nian Yang ; Chung-Kuang Chou ; Jihg-Hong Lin ; Ming-Syan Chen</i>	
DEVELOPER TOOLCHAINS FOR LARGE-SCALE ANALYTICS: TWO CASE STUDIES	1311
<i>Rosenthal, S. ; McMillan, S. ; Gaston, M.E.</i>	
ENTERPRISE SUBSCRIPTION CHURN PREDICTION	1317
<i>Vadakattu, R. ; Panda, B. ; Narayan, S. ; Godhia, H.</i>	
DATA DEIDENTIFICATION IN MEDICAL TRANSCRIPTIONS USING REGULAR EXPRESSIONS AND MACHINE LEARNING	1322
<i>Seeger, J. ; Culotta, A. ; Keller, J. ; van Kessel, P. ; Jugovich, M.</i>	
MACAU: LARGE-SCALE SKILL SENSE DISAMBIGUATION IN THE ONLINE RECRUITMENT DOMAIN	1324
<i>Qinlong Luo ; Meng Zhao ; Javed, F. ; Jacob, F.</i>	
GENOMIC ANALYSIS WITH MAPREDUCE	1330
<i>Wei Yi Liu ; Hui-I Hsiao ; Shih Yao Dai</i>	
EAGLE: USER PROFILE-BASED ANOMALY DETECTION FOR SECURING HADOOP CLUSTERS.....	1336
<i>Gupta, Chaitali ; Sinha, Ranjan ; Zhang, Yong</i>	
INVESTIGATING INSURANCE FRAUD USING SOCIAL MEDIA	1344
<i>Diaz-Granados, M. ; Diaz-Montes, J. ; Parashar, M.</i>	
A DOCUMENT-BASED DATA MODEL FOR LARGE SCALE COMPUTATIONAL MARITIME SITUATIONAL AWARENESS	1350
<i>Cazzanti, L. ; Millefiori, L.M. ; Arcieri, G.</i>	
MODELING SOCIAL INFLUENCES FROM CALL RECORDS AND MOBILE WEB BROWSING HISTORIES	1357
<i>Jhao-Yin Li ; Mi-Yen Yeh ; Ming-Syan Chen ; Jihg-Hong Lin</i>	
NEXT GENERATION BIOBANKS	1362
<i>Seebode, C. ; Ort, M. ; Hufnagl, P. ; Regenbrecht, C.R.A.</i>	
BUSINESS UNDERSTANDING, CHALLENGES AND ISSUES OF BIG DATA ANALYTICS FOR THE SERVICITIZATION OF A CAPITAL EQUIPMENT MANUFACTURER	1368
<i>Nino, Mikel ; Blanco, Jose Miguel ; Illarramendi, Arantza</i>	
DATA DRIVEN PREDICTIVE ANALYTICS FOR A SPINDLE'S HEALTH.....	1378
<i>Sardana, D. ; Bhatnagar, R. ; Pavel, R. ; Iverson, J.</i>	
A "SMART COMPONENT" DATA MODEL IN PLM.....	1388
<i>Yunpeng Li ; Roy, U. ; Seung-Jun Shin ; Lee, Y.T.</i>	

BIG DATA PROCESS ANALYTICS FOR CONTINUOUS PROCESS IMPROVEMENT IN MANUFACTURING.....	1398
<i>Stojanovic, N. ; Dinic, M. ; Stojanovic, L.</i>	
AUTOMATED UNCERTAINTY QUANTIFICATION ANALYSIS USING A SYSTEM MODEL AND DATA	1408
<i>Nannapaneni, S. ; Mahadevan, S. ; Lechevalier, D. ; Narayanan, A. ; Rachuri, S.</i>	
ANALYSIS AND OPTIMIZATION IN SMART MANUFACTURING BASED ON A REUSABLE KNOWLEDGE BASE FOR PROCESS PERFORMANCE MODELS	1418
<i>Brodsky, A. ; Guodong Shao ; Krishnamoorthy, M. ; Narayanan, A. ; Menasce, D. ; Ak, R.</i>	
A NEURAL NETWORK META-MODEL AND ITS APPLICATION FOR MANUFACTURING	1428
<i>Lechevalier, D. ; Hudak, S. ; Ak, R. ; Lee, Y.T. ; Fofou, S.</i>	
PERFORMANCE ASSESSMENT AND UNCERTAINTY QUANTIFICATION OF PREDICTIVE MODELS FOR SMART MANUFACTURING SYSTEMS.....	1436
<i>Oneto, L. ; Orlandi, I. ; Anguita, D.</i>	
TIME COMPLEXITY AND ARCHITECTURE OF A CLOUD BASED PROGNOSTICS SYSTEM FOR A MULTI-CLIENT CONDITION MONITORING ACTIVITY	1446
<i>Natarajan, Ashwin K.Thillai ; Kamarthi, Sagar</i>	
REAL-TIME ENERGY PREDICTION FOR A MILLING MACHINE TOOL USING SPARSE GAUSSIAN PROCESS REGRESSION	1451
<i>Park, J. ; Law, K.H. ; Bhinge, R. ; Chen, M. ; Dornfeld, D. ; Rachuri, S.</i>	
PARALLEL PARTICLE SWARM OPTIMIZATION (PPSO) CLUSTERING FOR LEARNING ANALYTICS.....	1461
<i>Govindarajan, K. ; Boulanger, D. ; Kumar, V.S. ; Kinshuk</i>	
ANALYSIS AND PREDICTION OF E-CUSTOMERS' BEHAVIOR BY MINING CLICKSTREAM DATA.....	1466
<i>Silahtaroglu, G. ; Donertasli, H.</i>	
HIGH QUALITY CLUSTERING OF BIG DATA AND SOLVING EMPTY-CLUSTERING PROBLEM WITH AN EVOLUTIONARY HYBRID ALGORITHM	1473
<i>Karimov, J. ; Ozbayoglu, M.</i>	
AGILE TEXT MINING WITH SHERLOK	1479
<i>Richardet, R. ; Chappelier, J.-C. ; Tripathy, S. ; Hill, S.</i>	
SCALABLE ADAPTIVE LABEL PROPAGATION IN GRAPPA	1485
<i>Farnadi, G. ; Mahdavifar, Z. ; Keller, I. ; Nelson, J. ; Teredesai, A. ; Moens, M.-F. ; De Cock, M.</i>	
PROFILING SUBSCRIBERS ACCORDING TO THEIR INTERNET USAGE CHARACTERISTICS AND BEHAVIORS	1492
<i>Oztoprak, K.</i>	
QUERIE RELOADED: USING MATRIX FACTORIZATION TO IMPROVE DATABASE QUERY RECOMMENDATIONS.....	1500
<i>Eirinaki, M. ; Patel, S.</i>	
MONITORING ADOLESCENT ALCOHOL USE VIA MULTIMODAL ANALYSIS IN SOCIAL MULTIMEDIA	1509
<i>Ran Pang ; Baretto, A. ; Kautz, H. ; Jiebo Luo</i>	
AN EFFICIENT MAP-REDUCE ALGORITHM FOR COMPUTING FORMAL CONCEPTS FROM BINARY DATA	1519
<i>Bhatnagar, R. ; Kumar, L.</i>	
LEARNING RELAXED 3-CLUSTERS FROM PAIRS OF RELATED DATASETS	1529
<i>Patchala, J. ; Bhatnagar, R.</i>	
PARALLEL INFORMATION FUSION METHOD FOR MICROARRAY DATA ANALYSIS	1539
<i>Jun Meng ; Rui Li ; Jing Zhang</i>	
A-STAR ALGORITHM BASED ON-DEMAND ROUTING PROTOCOL FOR HIERARCHICAL LEO/MEO SATELLITE NETWORKS	1545
<i>Xuezhi Ji ; Lixiang Liu ; Pei Zhao ; Dapeng Wang</i>	
GRANULAR MODELING WITH FUZZY COMPARATORS.....	1550
<i>Sosnowski, L. ; Szczuka, M. ; Slezak, D.</i>	
AGGLOMERATIVE ALGORITHM TO DISCOVER SEMANTICS FROM UNSTRUCTURED BIG DATA.....	1556
<i>I-Jen Chiang</i>	
A GRANULAR APPROACH FOR IDENTIFYING USER KNOWLEDGE.....	1564
<i>Denzler, A. ; Wehrle, M. ; Meier, A.</i>	
TWITTER OPINION MINING FOR ADVERSE DRUG REACTIONS	1570
<i>Liang Wu ; Teng-Sheng Moh ; Khuri, N.</i>	

DATA DECOMPOSITION AND DUAL CLUSTERING FOR CLINICAL CARE MANAGEMENT	1575
<i>Tsumoto, S. ; Hirano, S. ; Iwata, H.</i>	
HOLISTIC ENTITY MATCHING ACROSS KNOWLEDGE GRAPHS	1585
<i>Pershina, M. ; Yakout, M. ; Chakrabarti, K.</i>	
GRC-BASED STATISTIC OPTIMIZATION ALGORITHM FOR BIG TRUTH TABLE	1591
<i>Chen Ze-hua ; Ma He ; Zhang Yu</i>	
MINING INCOMPLETE DATA WITH MANY ATTRIBUTE-CONCEPT VALUES AND "DO NOT CARE" CONDITIONS	1597
<i>Clark, P.G. ; Grzymala-Busse, J.W.</i>	
CHINESE WALL SECURITY POLICIES INFORMATION FLOWS IN BUSINESS CLOUD	1603
<i>Tsau-Young Lin</i>	
GRANULAR FORMALIZATION OF MEDICAL DIAGNOSTIC PROCESS	1608
<i>Tsumoto, S. ; Hirano, S.</i>	
MOBILE GESTURE-BASED IPHONE USER AUTHENTICATION	1615
<i>Khare, K. ; Teng-Sheng Moh</i>	
COST AND DATA EXPLORATION CONSIDERATIONS FOR BIG DATA PREDICTION ON THE CLOUD	1622
<i>Tseng, C. ; Tien Nguyen ; Sharma, C.</i>	
MINING LOCAL GAZETTEERS OF LITERARY CHINESE WITH CRF AND PATTERN BASED METHODS FOR BIOGRAPHICAL INFORMATION IN CHINESE HISTORY	1629
<i>Chao-Lin Liu ; Chih-Kai Huang ; Hongsu Wang ; Bol, P.K.</i>	
TOWARDS A MOBILE SOCIAL DATA COMMONS	1639
<i>Greenway, G. ; Mack, L. ; Blanke, T. ; Cote, M. ; Heath, T.</i>	
SCALING OUT FOR EXTREME SCALE CORPUS DATA	1643
<i>Coole, M. ; Rayson, P. ; Mariani, J.</i>	
METAPHOR MINING IN HISTORICAL GERMAN NOVELS: AN UNSUPERVISED LEARNING APPROACH	1650
<i>Pernes, S.</i>	
PREDICTING SOCIAL TRENDS FROM NON-PHOTOGRAPHIC IMAGES ON TWITTER	1653
<i>Yazdani, M. ; Manovich, L.</i>	
THE CODING OF LITERARY FORM: DATA MINING AND THE INFORMATION STRUCTURE OF HISTORICAL TEXTS	1661
<i>Liddle, D.</i>	
PLOT ARCHEOLOGY: A VECTOR-SPACE MODEL OF NARRATIVE STRUCTURE	1667
<i>Schmidt, B.M.</i>	
A METHOD FOR CROSS-DOCUMENT NARRATIVE ALIGNMENT OF A TWO-HUNDRED-SIXTY-MILLION WORD CORPUS	1673
<i>Miller, Ben ; Olive, Jennifer ; Gopavaram, Shakthidhar ; Zhao, Yanjun ; Shrestha, Ayush ; Berger, Cynthia</i>	
MIXED-INITIATIVE SOCIAL MEDIA ANALYTICS AT THE WORLD BANK: OBSERVATIONS OF CITIZEN SENTIMENT IN TWITTER DATA TO EXPLORE "TRUST" OF POLITICAL ACTORS AND STATE INSTITUTIONS AND ITS RELATIONSHIP TO SOCIAL PROTEST	1678
<i>Calderon, N.A. ; Fisher, B. ; Hemsley, J. ; Ceskavich, B. ; Jansen, G. ; Marciano, R. ; Lemieux, V.L.</i>	
WORKLOAD-DRIVEN ADAPTIVE DATA PARTITIONING AND DISTRIBUTION - THE CUMULUS APPROACH	1688
<i>Fetai, I. ; Murrezzan, D. ; Schuldt, H.</i>	
ACCOUNT CLUSTERING IN MULTI-TENANT STORAGE MANAGEMENT ENVIRONMENTS	1698
<i>Madl, Gabor ; Routray, Ramani ; Song, Yang ; Jain, Rakesh</i>	
FINE-TUNING THE CONSISTENCY-LATENCY TRADE-OFF IN QUORUM-REPLICATED DISTRIBUTED STORAGE SYSTEMS	1708
<i>McKenzie, M. ; Hua Fan ; Golab, W.</i>	
PRIORITY REGISTER: APPLICATION-DEFINED REPLACEMENT ORDERINGS FOR AD HOC RECONCILIATION	1718
<i>Kumar, S.P. ; Lefebvre, S. ; Minyoung Kim ; Stehr, M.O.</i>	
A GENERALIZED FLOW FOR MULTI-CLASS AND BINARY CLASSIFICATION TASKS: AN AZURE ML APPROACH	1728
<i>Bihis, M. ; Roychowdhury, S.</i>	
COMPARISON OF EAGER AND QUORUM-BASED REPLICATION IN A CLOUD ENVIRONMENT	1738
<i>Stiemer, A. ; Fetai, I. ; Schuldt, H.</i>	
TOWARDS A TAXONOMY OF STANDARDS IN SMART DATA	1749
<i>Lenk, A. ; Bonorden, L. ; Hellmanns, A. ; Roedder, N. ; Jaehnichen, S.</i>	

MARLIN: TAMING THE BIG STREAMING DATA IN LARGE SCALE VIDEO SIMILARITY SEARCH	1755
<i>Nan Zhu ; Wenbo He ; Yu Hua ; Yixin Chen</i>	
INDEXING HISTORICAL SPATIO-TEMPORAL DATA IN THE CLOUD	1765
<i>Chong Zhang ; Xiaoying Chen ; Bin Ge ; Weidong Xiao</i>	
PUSH-BASED SYSTEM FOR MOLECULAR SIMULATION DATA ANALYSIS	1775
<i>Grupcev, V. ; Yi-Cheng Tu ; Fogarty, J. ; Pandit, S.</i>	
CHALLENGES AND OPPORTUNITIES ON NETWORK RESOURCE MANAGEMENT IN DCN WITH SDN	1785
<i>Guan Xu ; Jun Yang ; Bin Dai</i>	
ON THE IMPLEMENTATION OF ZIGZAG CODES FOR DISTRIBUTED STORAGE SYSTEM	1791
<i>Lijia Lu ; Hui Li ; Jun Chen ; Bing Zhu ; Weijuan Yin</i>	
A COMPREHENSIVE EVALUATION OF NOSQL DATASTORES IN THE CONTEXT OF HISTORIANS AND SENSOR DATA ANALYSIS	1797
<i>Kalakanti, A.K. ; Sudhakaran, V. ; Raveendran, V. ; Menon, N.</i>	
LEARNING CLASSIFIERS FROM REMOTE RDF DATA STORES AUGMENTED WITH RDFS SUBCLASS HIERARCHIES	1807
<i>Lin, H.T. ; Ngot Bui ; Honavar, V.</i>	
DISTINGER: A DISTRIBUTED GRAPH DATA STRUCTURE FOR MASSIVE DYNAMIC GRAPH PROCESSING	1814
<i>Guoyao Feng ; Xiao Meng ; Ammar, K.</i>	
LITEMAT: A SCALABLE, COST-EFFICIENT INFERENCE ENCODING SCHEME FOR LARGE RDF GRAPHS	1823
<i>Cure, Olivier ; Naacke, Hubert ; Randriamalala, Tendry ; Amann, Bernd</i>	
MQUERY: A QUERY LANGUAGE FOR SCIENTIFIC MESHES	1831
<i>Mahdiraji, A.R. ; Baumann, P.</i>	
A FAST PARALLEL ALGORITHM FOR COUNTING TRIANGLES IN GRAPHS USING DYNAMIC LOAD BALANCING	1839
<i>Arifuzzaman, Shaikh ; Khan, Maleq ; Marathe, Madhav</i>	
SCALABLE STORAGE STRUCTURE FOR PATTERN MATCHING ON BIG GRAPH DATA	1848
<i>Balaji, J. ; Sunderraman, R.</i>	
EMPLOYING IN-MEMORY DATA GRIDS FOR DISTRIBUTED GRAPH PROCESSING	1856
<i>Tasci, S. ; Demirbas, M.</i>	
CURRENT SECURITY THREATS AND PREVENTION MEASURES RELATING TO CLOUD SERVICES, HADOOP CONCURRENT PROCESSING, AND BIG DATA	1865
<i>Sharif, A. ; Cooney, S. ; Shengqi Gong ; Vitek, D.</i>	
SECURITY FOR THE SCIENTIFIC DATA SERVICES FRAMEWORK	1871
<i>Jinoh Kim ; Bin Dong ; Byna, S. ; Kesheng Wu</i>	
A NOVEL FRAMEWORK FOR MITIGATING INSIDER ATTACKS IN BIG DATA SYSTEMS	1876
<i>Aditham, S. ; Ranganathan, N.</i>	
HETEROGENEOUS K-ANONYMIZATION WITH HIGH UTILITY	1886
<i>Doka, K. ; Mingqiang Xue ; Tsoumakos, D. ; Karras, P. ; Cuzzocrea, A. ; Koziris, N.</i>	
MULTI-PROBE RANDOM PROJECTION CLUSTERING TO SECURE VERY LARGE DISTRIBUTED DATASETS	1891
<i>Carraher, L.A. ; Wilsey, P.A. ; Moitra, A. ; Dey, S.</i>	
FAST SUMMARIZATION AND ANONYMIZATION OF MULTIVARIATE BIG TIME SERIES	1901
<i>Ruta, D. ; Ling Cen ; Damiani, E.</i>	
TOWARD BIG DATA RISK ANALYSIS	1905
<i>Damiani, E.</i>	
A DISTRIBUTED FRAMEWORK FOR SUPPORTING ADAPTIVE ENSEMBLE-BASED INTRUSION DETECTION	1910
<i>Cuzzocrea, A. ; Folino, G. ; Sabatino, P.</i>	
SIMPLIFYING WEB ANALYTICS FOR DIGITAL MARKETING	1917
<i>Bengel, A. ; Shawki, A. ; Aggarwal, D.</i>	
PAUSE: A PRIVACY ARCHITECTURE FOR HETEROGENEOUS BIG DATA ENVIRONMENTS	1919
<i>Jutla, D.N. ; Bodorik, P.</i>	
SPATIO-TEMPORAL QUERIES IN HBASE	1929
<i>Xiaoying Chen ; Chong Zhang ; Bin Ge ; Weidong Xiao</i>	
COMPONENT BASED DATAFLOW PROCESSING FRAMEWORK	1938
<i>Gyurjyan, V. ; Bartle, A. ; Lukashin, C. ; Mancilla, S. ; Oyarzun, R. ; Vakhnin, A.</i>	
EARTH SCIENCE DATA FUSION WITH EVENT BUILDING APPROACH	1943
<i>Lukashin, C. ; Bartle, A. ; Callaway, E. ; Gyirjyan, V. ; Mancilla, S. ; Oyarzun, R. ; Vakhnin, A.</i>	

CLIMATE MODEL DIAGNOSTIC ANALYZER	1948
<i>Seungwon Lee ; Lei Pan ; Chengxing Zhai ; Benyang Tang ; Kubar, T. ; Jia Zhang ; Wei Wang</i>	
HIGH PERFORMANCE ANALYSIS OF BIG SPATIAL DATA	1953
<i>Haynes, D. ; Ray, S. ; Manson, S.M. ; Soni, A.</i>	
INTERNATIONAL STANDARD "OGC® MOVING FEATURES" TO ADDRESS "4VS" ON LOCATIONAL BIGDATA	1958
<i>Asahara, A. ; Hayashi, H. ; Ishimaru, N. ; Shibasaki, R. ; Kanasugi, H.</i>	
OPTIMIZING APACHE NUTCH FOR DOMAIN SPECIFIC CRAWLING AT LARGE SCALE	1967
<i>Lopez, L.A. ; Duerr, R. ; Khalsa, S.J.S.</i>	
A HADOOP-BASED VISUALIZATION AND DIAGNOSIS FRAMEWORK FOR EARTH SCIENCE DATA	1972
<i>Shujia Zhou ; Xi Yang ; Li, X. ; Matsui, T. ; Si Liu ; Xian-He Sun ; Tao, W.</i>	
ENABLING SCIENTIFIC DATA STORAGE AND PROCESSING ON BIG-DATA SYSTEMS	1978
<i>Biokaghazadeh, S. ; Yiqi Xu ; Zhou, S. ; Ming Zhao</i>	
LIGHT-WEIGHT PARALLEL PYTHON TOOLS FOR EARTH SYSTEM MODELING WORKFLOWS	1985
<i>Paul, K. ; Mickelson, S. ; Dennis, J.M. ; Haiying Xu ; Brown, D.</i>	
WDCLOUD: AN END TO END SYSTEM FOR LARGE-SCALE WATERSHED DELINEATION ON CLOUD	1995
<i>In Kee Kim ; Steele, J. ; Castronova, A.M. ; Goodall, J.L. ; Humphrey, M.</i>	
INTEGRATING 'BIG' GEOSCIENCE DATA INTO THE PETASCALE NATIONAL ENVIRONMENTAL RESEARCH INTEROPERABILITY PLATFORM (NERDIP): SUCCESSES AND UNFORESEEN CHALLENGES	2005
<i>Wyborn, L. ; Evans, B.J.K.</i>	
AN OPTIMIZED INTERESTINGNESS HOTSPOT DISCOVERY FRAMEWORK FOR LARGE GRIDDED SPATIO-TEMPORAL DATASETS	2010
<i>Akdag, F. ; Eick, C.F.</i>	
SCISPARK: APPLYING IN-MEMORY DISTRIBUTED COMPUTING TO WEATHER EVENT DETECTION AND TRACKING	2020
<i>Palamuttam, R. ; Mogrovejo, R.M. ; Mattmann, C. ; Wilson, B. ; Whitehall, K. ; Verma, R. ; McGibbney, L. ; Ramirez, P.</i>	
DETECTING ENVIRONMENTAL DISASTERS IN DIGITAL NEWS ARCHIVES	2027
<i>Yzaguirre, A. ; Warren, R. ; Smit, M.</i>	
IS APACHE SPARK SCALABLE TO SEISMIC DATA ANALYTICS AND COMPUTATIONS?	2036
<i>Yuzhong Yan ; Lei Huang ; Liqi Yi</i>	
ON THE EFFICIENT EVALUATION OF ARRAY JOINS	2046
<i>Baumann, P. ; Merticariu, V.</i>	
BUSINESS INFORMATION MODELING: A METHODOLOGY FOR DATA-INTENSIVE PROJECTS, DATA SCIENCE AND BIG DATA GOVERNANCE	2056
<i>Priebe, T. ; Markus, S.</i>	
THE NEED FOR NEW PROCESSES, METHODOLOGIES AND TOOLS TO SUPPORT BIG DATA TEAMS AND IMPROVE BIG DATA PROJECT EFFECTIVENESS	2066
<i>Saltz, J.S.</i>	
TOWARDS METHODS FOR SYSTEMATIC RESEARCH ON BIG DATA	2072
<i>Das, M. ; Renhao Cui ; Campbell, D.R. ; Agrawal, G. ; Ramnath, R.</i>	
TOWARDS A BIG DATA THEORY MODEL	2082
<i>Pospiech, M. ; Felden, C.</i>	
THREE CRITICAL MATTERS IN BIG DATA PROJECTS FOR E-SCIENCE: DIFFERENT USER GROUPS, THE MUTUALLY CONSTITUTIVE PERSPECTIVE, AND VIRTUAL ORGANIZATIONAL CAPACITY	2091
<i>Kee, K.F.</i>	
EXPLORING THE PROCESS OF DOING DATA SCIENCE VIA AN ETHNOGRAPHIC STUDY OF A MEDIA ADVERTISING COMPANY	2098
<i>Saltz, J.S. ; Shamshurin, I.</i>	
FORECAST UPC-LEVEL FMCG DEMAND, PART I: EXPLORATORY ANALYSIS AND VISUALIZATION	2106
<i>DaZhi Yang ; Goh, G.S.W. ; Chi Xu ; Zhang, A.N. ; Akcan, O.</i>	
FORECAST UPC-LEVEL FMCG DEMAND, PART II: HIERARCHICAL RECONCILIATION	2113
<i>DaZhi Yang ; Goh, G.S.W. ; Siwei Jiang ; Zhang, A.N. ; Akcan, O.</i>	
SPARSITY ADJUSTED INFORMATION GAIN FOR FEATURE SELECTION IN SENTIMENT ANALYSIS	2122
<i>Ong, B.Y. ; Goh, S.W. ; Chi Xu</i>	

DYNAMIC AGGREGATION FOR TIME SERIES FORECASTING	2129
<i>Iosevich, S. ; Arutyunyan, G. ; Hou, Z.</i>	
BIG DATA ANALYTICS FOR EMPOWERING MILK YIELD PREDICTION IN DAIRY SUPPLY CHAINS	2132
<i>Yan, W.J. ; Chen, X. ; Akcan, O. ; Lim, J. ; Yang, D.</i>	
PROFIT ESTIMATION ERROR ANALYSIS IN RECOMMENDER SYSTEMS BASED ON ASSOCIATION RULES	2138
<i>Ertek, G. ; Xu Chi ; Yee, G. ; Ong Boon Yong ; Byung-Geun Choi</i>	
GRAPH-BASED ANALYSIS OF RESOURCE DEPENDENCIES IN PROJECT NETWORKS	2143
<i>Ertek, G. ; Byung-Geun Choi ; Xu Chi ; DaZhi Yang ; Ong Boon Yong</i>	
A DATA FUSION FRAMEWORK FOR LARGE-SCALE MEASUREMENT PLATFORMS	2150
<i>Rattadilok, P. ; McCall, J. ; Burbridge, T. ; Soppera, A. ; Eardley, P.</i>	
SENSOR EVENT MINING WITH HYBRID ENSEMBLE LEARNING AND EVOLUTIONARY FEATURE SUBSET SELECTION MODEL	2159
<i>Mehdiyev, N. ; Krumeich, J. ; Werth, D. ; Loos, P.</i>	
OPTIMIZATION OF SYSTEM ARCHITECTURE FOR BIG DATA ANALYSIS IN CLIMATE SCIENCE	2169
<i>Huikyo Lee ; Cinquini, L. ; Crichton, D. ; Braverman, A.</i>	
IN-SITU ANALYTICS FOR TOMOGRAPHIC IMAGING IN SENSOR NETWORK	2173
<i>Kamath, G. ; Wen-Zhan Song</i>	
ONTOLOGY-DRIVE DATA ACCESS AT THE NASA EARTH EXCHANGE	2177
<i>Huffer, B. ; Cotnoir, M. ; Gleason, J.</i>	
STRATEGIE ROADMAP FOR THE EARTH SYSTEM GRID FEDERATION	2182
<i>Williams, D.N. ; Lautenschlager, M. ; Balaji, V. ; Cinquini, L. ; DeLuca, C. ; Denvil, S. ; Duffy, D. ; Evans, B. ; Ferraro, R. ; Juckes, M. ; Trenham, C.</i>	
CONSTRAINED REGION SELECTION METHOD BASED ON CONFIGURATION SPACE FOR VISUALIZATION IN SCIENTIFIC DATASET SEARCH	2191
<i>Takeuchi, S. ; Sugiura, K. ; Akahoshi, Y. ; Zettsu, K.</i>	
ENHANCING SCIENCE SUPPORT IN SQL	2201
<i>Baumann, P. ; Misev, D.</i>	
MODELING COMMUNITY DETECTION USING SLOW MIXING RANDOM WALKS	2205
<i>Torghabeh, R.P. ; Santhanam, N.P.</i>	
DIMENSIONAL SCALABILITY OF SUPERVISED AND UNSUPERVISED CONCEPT DRIFT DETECTION: AN EMPIRICAL STUDY	2212
<i>Lavaire, J.D.D. ; Singh, A. ; Yousef, M. ; Singh, S. ; Xiaodong Yue</i>	
EFFICIENT CHANGE DETECTION FOR HIGH DIMENSIONAL DATA STREAMS	2219
<i>Georgakopoulos, Spiros V. ; Tasoulis, Sotiris K. ; Plagianakos, Vassilis P.</i>	
BIG DATA ANALYTICS FOR DEMAND RESPONSE: CLUSTERING OVER SPACE AND TIME	2223
<i>Chelmis, C. ; Kolte, J. ; Prasanna, V.K.</i>	
FINDING BANDED PATTERNS IN BIG DATA USING SAMPLING	2233
<i>Abdullahi, F.B. ; Coenen, F. ; Martin, R.</i>	
SCALABLE PREFERENCE QUERIES FOR HIGH-DIMENSIONAL DATA USING MAP-REDUCE	2243
<i>Guzun, G. ; Tosado, J.E. ; Canahuate, G.</i>	
DISCOVERING TIME-EVOLVING INFLUENCE FROM DYNAMIC HETEROGENEOUS GRAPHS	2253
<i>Chuan Hu ; Huiping Cao</i>	
COMBINING ACTIVITY-EVALUATION INFORMATION WITH NMF FOR TRUST-LINK PREDICTION IN SOCIAL MEDIA	2263
<i>Matsutani, K. ; Kumano, M. ; Kimura, M. ; Saito, K. ; Ohara, K. ; Motoda, H.</i>	
IDENTIFYING ACTIONABLE MESSAGES ON SOCIAL MEDIA	2273
<i>Spasojevic, N. ; Rao, A.</i>	
KLOUT SCORE: MEASURING INFLUENCE ACROSS MULTIPLE SOCIAL NETWORKS	2282
<i>Rao, A. ; Spasojevic, N. ; Zhisheng Li ; Dsouza, T.</i>	
TOP (K_1, K_2) DISTANCE-BASED OUTLIERS DETECTION IN AN UNCERTAIN DATASET	2290
<i>Liu, Fei ; Jia, Yan</i>	
UNDERSTANDING THE TIME CHARACTERISTIC OF USER BEHAVIOR ON ONLINE FORUMS	2300
<i>Guirong Chen ; Ning Wang ; Fengqin Zhang ; Hua Jiang</i>	
CHARACTERIZING SUPER SPREADING IN MICROBLOG: AN EPIDEMIC-BASED MODEL	2307
<i>Yu Liu ; Bin Wu ; Bai Wang</i>	

A COMMUNITY DETECTION METHOD BASED ON K-SHELL.....	2314
<i>Yang Wang ; Liutong Xu ; Bin Wu</i>	
HOW MUCH IS YOUR INFORMATION WORTH - A METHOD FOR REVENUE GENERATION FOR YOUR INFORMATION.....	2320
<i>Divya Rao ; Wee Keong Ng</i>	
EFFICIENT LARGE SCALE DISTRIBUTED MATRIX COMPUTATION WITH SPARK.....	2327
<i>Rong Gu ; Yun Tang ; Zhaokang Wang ; Shuai Wang ; Xusen Yin ; Chunfeng Yuan ; Yihua Huang</i>	
A COLLABORATIVE FILTERING ALGORITHM FUSING USER-BASED, ITEM-BASED AND SOCIAL NETWORKS	2337
<i>Bailing Wang ; Junheng Huang ; Libing Ou ; Rui Wang</i>	
MINING THE RELATION BETWEEN DORM ARRANGEMENT AND STUDENT PERFORMANCE.....	2344
<i>Man Li ; Ruisheng Shi</i>	
A PROACTIVE DISCOVERY AND FILTERING SOLUTION ON PHISHING WEBSITES	2348
<i>lv Fang ; Wang Bailing ; Huang Junheng ; Sun Yushan ; Wei Yuliang</i>	
FINDING COMMUNITY STRUCTURE VIA ROUGH K-MEANS IN SOCIAL NETWORK	2356
<i>Yunlei Zhang ; Bin Wu</i>	
A SURVEY OF SEMANTIC SIMILARITY AND ITS APPLICATION TO SOCIAL NETWORK ANALYSIS	2362
<i>Shuang Zhang ; Xuefeng Zheng ; Changjun Hu</i>	
DYNAMIC COMMUNITY DETECTION BASED ON GAME THEORY IN SOCIAL NETWORKS	2368
<i>Fei Jiang ; Jin Xu</i>	
THE VALUE OF ANALYTICAL QUERIES ON SOCIAL NETWORKS.....	2374
<i>de Rougemont, M. ; Vimont, G.</i>	
A COLLABORATIVE FILTERING ALGORITHM BASED ON SOCIAL NETWORK INFORMATION	2384
<i>Rui Wang ; Bailing Wang ; Junheng Huang</i>	
EFFICIENT APPROXIMATION ALGORITHMS TO DETERMINE MINIMUM PARTIAL DOMINATING SETS IN SOCIAL NETWORKS	2390
<i>Campan, A. ; Truta, T.M. ; Beckerich, M.</i>	
TIES THAT MATTER	2398
<i>Chowdhary, G. ; Bandyopadhyay, S.</i>	
SENTIMENT EXPRESSION VIA EMOTICONS ON SOCIAL MEDIA.....	2404
<i>Hao Wang ; Castanon, J.A.</i>	
ON COMPRESSING MASSIVE STREAMING GRAPHS WITH QUADTREES	2409
<i>Nelson, M. ; Radhakrishnan, S. ; Chatterjee, A. ; Sekharan, C.N.</i>	
SOCIAL SET VISUALIZER: A SET THEORETICAL APPROACH TO BIG SOCIAL DATA ANALYTICS OF REAL-WORLD EVENTS.....	2418
<i>Flesch, B. ; Vatrappu, R. ; Mukkamala, R.R. ; Hussain, A.</i>	
A NOVEL SYMBOLIZATION TECHNIQUE FOR TIME-SERIES OUTLIER DETECTION.....	2428
<i>Smith, G. ; Goulding, J.</i>	
VOLATILITY MATRIX INFERENCE IN HIGH-FREQUENCY FINANCE WITH REGULARIZATION AND EFFICIENT COMPUTATIONS	2437
<i>Jian Zou ; Yunbo An ; Hong Yan</i>	
SHAPING DATA: VISUALIZATION UNDER CONSTRUCTION	2445
<i>Bieh-Zimmert, O. ; Felden, C.</i>	
IMMERSIVE VISUALIZATION FOR MATERIALS SCIENCE DATA ANALYSIS USING THE OCULUS RIFT	2453
<i>Drouhard, M. ; Steed, C.A. ; Hahn, S. ; Proffen, T. ; Daniel, J. ; Matheson, M.</i>	
SPATIO-TEMPORAL SIMILARITY SEARCH METHOD FOR DISASTER ESTIMATION.....	2462
<i>Hayashi, H. ; Asahara, A. ; Sugaya, N. ; Ogawa, Y. ; Tomita, H.</i>	
SCALABLE DENTAL COMPUTING ON CYBERINFRASTRUCTURE.....	2470
<i>Hui Zhang ; Riqing Chen ; Guangchen Ruan ; Ando, M.</i>	
WRANGLER'S USER ENVIRONMENT: A SOFTWARE FRAMEWORK FOR MANAGEMENT OF DATA-INTENSIVE COMPUTING SYSTEM.....	2479
<i>Jordan, C. ; Walling, D. ; Weijia Xu ; Mock, S.A. ; Gaffney, N. ; Stanzione, D.</i>	
VISUAL ANALYSIS OF LARGE-SCALE LIDAR POINT CLOUDS	2487
<i>Wanbo Luo ; Hui Zhang</i>	
A DATABASE-BASED DISTRIBUTED COMPUTATION ARCHITECTURE WITH ACCUMULO AND D4M: AN APPLICATION OF EIGENSOLVER FOR LARGE SPARSE MATRIX.....	2493
<i>Yin Huang ; Yesha, Y. ; Shujia Zhou</i>	

TEXTURE-BASED EDGE BUNDLING: A WEB-BASED APPROACH FOR INTERACTIVELY VISUALIZING LARGE GRAPHS	2501
<i>Jieting Wu ; Yu, L. ; Hongfeng Yu</i>	
BIG DATA PROVENANCE: CHALLENGES, STATE OF THE ART AND OPPORTUNITIES	2509
<i>Jianwu Wang ; Crawl, D. ; Purawat, S. ; Mai Nguyen ; Altintas, I.</i>	
PERFORMANCE EVALUATION OF ENABLING LOGISTIC REGRESSION FOR BIG DATA WITH R	2517
<i>Ruizhu Huang ; Weijia Xu</i>	
SKILL GROUPING METHOD: MINING AND CLUSTERING SKILL DIFFERENCES FROM BODY MOVEMENT BIGDATA	2525
<i>Yamagiwa, S. ; Kawahara, Y. ; Tabuchi, N. ; Watanabe, Y. ; Naruo, T.</i>	
REGULARIZED AND SPARSE STOCHASTIC K-MEANS FOR DISTRIBUTED LARGE-SCALE CLUSTERING	2535
<i>Jumutc, V. ; Langone, R. ; Suykens, J.A.K.</i>	
JOIN ALGORITHMS ON GPUS: A REVISIT AFTER SEVEN YEARS	2541
<i>Ran Rui ; Hao Li ; Yi-Cheng Tu</i>	
A DATA-DRIVEN APPROACH TOWARDS PATIENT IDENTIFICATION FOR TELEHEALTH PROGRAMS	2551
<i>Ganser, M. ; Dhar, S. ; Kurup, U. ; Cunha, C. ; Gacic, A.</i>	
ENSEMBLE PREDICTION OF VASCULAR INJURY IN TRAUMA CARE: INITIAL EFFORTS TOWARDS DATA-DRIVEN, LOW-COST SCREENING	2560
<i>Metzger, M. ; Howard, M. ; Kellogg, L. ; Kundi, R.</i>	
M-SEQ: EARLY DETECTION OF ANXIETY AND DEPRESSION VIA TEMPORAL ORDERS OF DIAGNOSES IN ELECTRONIC HEALTH DATA	2569
<i>Zhang, Jinghe ; Xiong, Haoyi ; Huang, Yu ; Wu, Hao ; Leach, Kevin ; Barnes, Laura E.</i>	
USING CLINICAL DATA, HYPOTHESIS GENERATION TOOLS AND PUBMED TRENDS TO DISCOVER THE ASSOCIATION BETWEEN DIABETIC RETINOPATHY AND ANTIHYPERTENSIVE DRUGS	2578
<i>Senter, K. ; Sukumar, S.R. ; Patton, R.M. ; Chaum, E.</i>	
ENABLING GRAPH APPLIANCE FOR GENOME ASSEMBLY	2583
<i>Singh, R. ; Graves, J.A. ; Lee, S. ; Sukumar, S.R. ; Shankar, M.</i>	
A FRAMEWORK FOR CONSENSUAL AND ONLINE PRIVACY PRESERVING RECORD LINKAGE IN REAL-TIME	2591
<i>Muller, D. ; Mau, S. ; Cvijikj, I.P.</i>	
A MEMORY CAPACITY MODEL FOR HIGH PERFORMING DATA-FILTERING APPLICATIONS IN SAMZA FRAMEWORK	2600
<i>Tao Feng ; Zhenyun Zhuang ; Yi Pan ; Ramachandra, H.</i>	
ROBUST AND DISTRIBUTED WEB-SCALE NEAR-DUP DOCUMENT CONFLATION IN MICROSOFT ACADEMIC SERVICE	2606
<i>Chieh-Han Wu ; Yang Song</i>	
EVALUATION OF DATA QUALITY OF MULTISITE ELECTRONIC HEALTH RECORD DATA FOR SECONDARY ANALYSIS	2612
<i>Nobles, A.L. ; Vilankar, K. ; Hao Wu ; Barnes, L.E.</i>	
CROWDMD: CROWDSOURCING-BASED APPROACH FOR DEDUPLICATION	2621
<i>Abboura, A. ; Sahrl, S. ; Ouziri, M. ; Benbernou, S.</i>	
DATA VERACITY ESTIMATION WITH ENSEMBLING TRUTH DISCOVERY METHODS	2628
<i>Berti-Equille, L.</i>	
DISTRIBUTED LIFE CYCLE SCHEDULING FOR CASCADED DATA PROCESSING	2637
<i>Sainik, L.</i>	
BIG DATA, BIG DATA QUALITY PROBLEM	2644
<i>Becker, D. ; McMullen, B. ; King, T.D.</i>	
DATA QUALITY ISSUES IN BIG DATA	2654
<i>Rao, D. ; Gudivada, V.N. ; Raghavan, V.V.</i>	
MACHINE LEARNING FOR STRESS DETECTION FROM ECG SIGNALS IN AUTOMOBILE DRIVERS	2661
<i>Keshan, N. ; Parimi, P.V. ; Bichindaritz, I.</i>	
SEQUENTIAL PATTERN MINING OF ELECTRONIC HEALTHCARE REIMBURSEMENT CLAIMS: EXPERIENCES AND CHALLENGES IN UNCOVERING HOW PATIENTS ARE TREATED BY PHYSICIANS	2670
<i>Malhotra, K. ; Hobson, T.C. ; Valkova, S. ; Pullum, L.L. ; Ramanathan, A.</i>	
SQL-LIKE BIG DATA ENVIRONMENTS: CASE STUDY IN CLINICAL TRIAL ANALYTICS	2680
<i>Grover, A. ; Gholap, J. ; Janeja, V.P. ; Yesha, Y. ; Chintalapati, R. ; Marwaha, H. ; Modi, K.</i>	

EXPLORING SPATIO-TEMPORAL-THEME CORRELATION BETWEEN PHYSICAL AND SOCIAL STREAMING DATA FOR EVENT DETECTION AND PATTERN INTERPRETATION FROM HETEROGENEOUS SENSORS.....	2690
<i>Minh-Son Dao ; Zettsu, K. ; Pongpaichet, S. ; Jalali, L. ; Jain, R.</i>	
MICRODATA ANALYSIS OF THE ACCOMMODATION SURVEY IN JAPANESE TOURISM STATISTICS.....	2700
<i>Sato, A.-H.</i>	
DETECTING RUMOR PATTERNS IN STREAMING SOCIAL MEDIA	2709
<i>Shihan Wang ; Terano, T.</i>	
A COLLABORATIVE FRAMEWORK FOR ANNOTATING ENERGY DATASETS	2716
<i>Hong-An Cao ; Kurniawan Wijaya, T. ; Aberer, K. ; Nunes, N.</i>	
THE RELATION BETWEEN FIRM AGE DISTRIBUTIONS AND THE DECAY RATE OF FIRM ACTIVITIES IN THE UNITED STATES AND JAPAN	2726
<i>Ishikawa, Atushi ; Fujimoto, Shouji ; Mizuno, Takayuki ; Watanabe, Tsutomu</i>	
AN EPIDEMIC SIMULATION WITH A DELAYED STOCHASTIC SIR MODEL BASED ON INTERNATIONAL SOCIOECONOMIC-TECHNOLOGICAL DATABASES	2732
<i>Sato, A.-H. ; Ito, I. ; Sawai, H. ; Iwata, K.</i>	
A SPATIO-TEMPORAL MULTIMEDIA BIG DATA FRAMEWORK FOR A LARGE CROWD.....	2742
<i>Sadiq, B. ; Reliman, F.U. ; Ahmad, A. ; Abdur Rahman, M. ; Ghani, S. ; Murad, A. ; Basalamah, S. ; Lbath, A.</i>	
DISTRIBUTED DYNAMIC ELASTIC NETS: A SCALABLE APPROACH FOR REGULARIZATION IN DYNAMIC MANUFACTURING ENVIRONMENTS	2752
<i>Ramakrishnan, N. ; Ghosh, R.</i>	
DIRECTIONAL DECISION LISTS.....	2762
<i>Goessling, M. ; Shan Kang</i>	
ANALYSIS OF KEY OPERATION PERFORMANCE DATA IN MANUFACTURING SYSTEMS.....	2767
<i>Ningxuan Kang ; Cong Zhao ; Jingshan Li ; Horst, J.A.</i>	
OUTLIER DETECTION FOR LARGE SCALE MANUFACTURING PROCESSES	2771
<i>Jauhri, A. ; McDanel, B. ; Connor, C.</i>	
FAST DETECTION OF MATERIAL DEFORMATION THROUGH STRUCTURAL DISSIMILARITY	2775
<i>Ushizima, Daniela ; Perciano, Talita ; Parkinson, Dilworth</i>	
DATA ANALYTICS AND UNCERTAINTY QUANTIFICATION FOR ENERGY PREDICTION IN MANUFACTURING.....	2782
<i>Ak, R. ; Bhing, R.</i>	
LAMBDA ARCHITECTURE FOR COST-EFFECTIVE BATCH AND SPEED BIG DATA PROCESSING	2785
<i>Kiran, M. ; Murphy, P. ; Monga, I. ; Dugan, J. ; Baveja, S.S.</i>	
NETWORK-AWARE RESOURCE MANAGEMENT FOR SCALABLE DATA ANALYTICS FRAMEWORKS	2793
<i>Renner, T. ; Thamsen, L. ; Kao, O.</i>	
ON A NEW APPROACH TO THE INDEX SELECTION PROBLEM USING MINING ALGORITHMS	2801
<i>Ameri, P. ; Meyer, J. ; Streit, A.</i>	
PREPARING, STORING, AND DISTRIBUTING MULTI-DIMENSIONAL SCIENTIFIC DATA	2811
<i>Devarakonda, R. ; Wei, Y. ; Thornton, M. ; Mayer, B. ; Thornton, P. ; Cook, B.</i>	
USE OF A METADATA DOCUMENTATION AND SEARCH TOOL FOR LARGE DATA VOLUMES: THE NGEA ARCTIC EXAMPLE.....	2814
<i>Devarakonda, R. ; Hook, L. ; Killeffer, T. ; Krassovski, M. ; Boden, T. ; Wullschleger, S.</i>	
DATA OPTIMISED COMPUTING FOR HETEROGENEOUS BIG DATA COMPUTING APPLICATIONS	2817
<i>Yang, E. ; Ross, D. ; Nagella, S. ; Turner, M. ; Kockelmann, W. ; Burca, G. ; Pouzols, F.M.</i>	
TOP-K COMPUTATIONS IN MAPREDUCE: A CASE STUDY ON RECOMMENDATIONS	2820
<i>Efthymiou, Vasilis ; Stefanidis, Kostas ; Ntoutsis, Eirini</i>	
A LSTM-BASED METHOD FOR STOCK RETURNS PREDICTION: A CASE STUDY OF CHINA STOCK MARKET	2823
<i>Kai Chen ; Yi Zhou ; Fangyan Dai</i>	
PREDICTING VARIOUS TYPES OF USER ATTRIBUTES IN TWITTER BY USING PERSONALIZED PAGERANK	2825
<i>Uesato, K. ; Asai, H. ; Yamana, H.</i>	
LARGE-SCALE LEARNING WITH ADAGRAD ON SPARK	2828
<i>Hadgu, A.T. ; Nigam, A. ; Diaz-Aviles, E.</i>	

PARALLELIZING NATURAL LANGUAGE TECHNIQUES FOR KNOWLEDGE EXTRACTION FROM CLOUD SERVICE LEVEL AGREEMENTS.....	2831
<i>Mittal, S. ; Joshi, K.P. ; Pearce, C. ; Joshi, A.</i>	
GRADIENT-BASED SIGNATURES FOR BIG MULTIMEDIA DATA.....	2834
<i>Beecks, C. ; Uysal, M.S. ; Seidl, T.</i>	
INDEXING MEDIA STORMS ON FLINK.....	2836
<i>Rafailidis, D. ; Antaris, S.</i>	
SCALING NLP ALGORITHMS TO MEET HIGH DEMAND.....	2839
<i>Stokes, C. ; Kumar, A. ; Choi, F. ; Weischedel, R.</i>	
THE NIST DATA SCIENCE EVALUATION SERIES: PART OF THE NIST INFORMATION ACCESS DIVISION DATA SCIENCE INITIATIVE.....	2840
<i>Dorr, B.J. ; Greenberg, C.S. ; Fontana, P. ; Przybocki, M. ; Le Bras, M. ; Ploehn, C. ; Aulov, O. ; Wo Chang</i>	
FLEXIBLE INGEST FRAMEWORK: A SCALABLE ARCHITECTURE FOR DYNAMIC ROUTING THROUGH COMPOSABLE PIPELINES.....	2843
<i>Samoylov, A. ; Schlachter, J.</i>	
A SCALABLE SOLUTION FOR GROUP FEATURE SELECTION.....	2846
<i>Govindan, P. ; Ruobing Chen ; Scheinberg, K. ; Srinivasan, S.</i>	
GENETIC DEEP NEURAL NETWORKS USING DIFFERENT ACTIVATION FUNCTIONS FOR FINANCIAL DATA MINING.....	2849
<i>Zhang, L.M.</i>	
PERFORMANCE OF GRAPH RECONSTRUCTION METHOD FOR LARGE-SCALE WEB GRAPH ANALYSIS.....	2852
<i>Takei, R. ; Niimi, A.</i>	
LOW LATENCY ANALYTICS FOR STREAMING TRAFFIC DATA WITH APACHE SPARK.....	2855
<i>Maarala, A.I. ; Rautiainen, M. ; Salmi, M. ; Pirttikangas, S. ; Riekkii, J.</i>	
HOW TO MAKE MONEY FROM YOUR INFORMATION AND KEEP YOUR PRIVACY.....	2859
<i>Rao, D. ; Wee Keong Ng</i>	
SCHEDULING OF BIG DATA APPLICATION WORKFLOWS IN CLOUD AND INTER-CLOUD ENVIRONMENTS.....	2862
<i>Rani, B.K. ; Babu, A.V.</i>	
PATIENT-LIKE-MINE: A REAL TIME, VISUAL ANALYTICS TOOL FOR CLINICAL DECISION SUPPORT.....	2865
<i>Li, P. ; Yates, S.N. ; Lovely, J.K. ; Larson, D.W.</i>	
A PRICING MECHANISM USING SOCIAL MEDIA AND WEB DATA TO INFER DYNAMIC CONSUMER VALUATIONS.....	2868
<i>Johnson, S.D. ; Kang-Yu Ni</i>	
EFFICIENT KEYWORD SEARCH ON GRAPHS USING MAPREDUCE.....	2871
<i>Yifan Hao ; Huiping Cao ; Yan Qi ; Chuan Hu ; Brahma, S. ; Jingyu Han</i>	
NON-BLOCKING ONE-PHASE COMMIT MADE POSSIBLE FOR DISTRIBUTED TRANSACTIONS OVER REPLICATED DATA.....	2874
<i>Yuqing Zhu</i>	
A LARGE SCALE EXAMINATION OF VEHICLE RECORDER DATA TO UNDERSTAND RELATIONSHIP BETWEEN DRIVERS' BEHAVIORS AND THEIR PAST DRIVING HISTORIES.....	2877
<i>Yokoyama, D. ; Toyoda, M.</i>	
ONLINE PATTERN MINING FOR HIGH-DIMENSIONAL DATA STREAMS.....	2880
<i>Yamamoto, Yoshitaka ; Iwanuma, Koji</i>	
MODELING THE LEARNING BEHAVIORS OF MASSIVE OPEN ONLINE COURSES.....	2883
<i>Zhenhui Liu ; Jingjing He ; Yufei Xue ; Zhenzhong Huang ; Manli Li ; Zhihui Du</i>	
DATA CONFIDENTIALITY CHALLENGES IN BIG DATA APPLICATIONS.....	2886
<i>Jian Yin ; Dongfang Zhao</i>	
FACTORIZATION MACHINES WITH FOLLOW-THE-REGULARIZED-LEADER FOR CTR PREDICTION IN DISPLAY ADVERTISING.....	2889
<i>Anh-Phuong Ta</i>	
TAXI TRIP TIME PREDICTION USING SIMILAR TRIPS AND ROAD NETWORK DATA.....	2892
<i>Singh, A.D. ; Wei Wu ; Shili Xiang ; Krishnaswamy, S.</i>	
USING WORD2VEC TO PROCESS BIG TEXT DATA.....	2895
<i>Long Ma ; Yanqing Zhang</i>	
INFERRING BIKE TRIP PATTERNS FROM BIKE SHARING SYSTEM OPEN DATA.....	2898
<i>Longbiao Chen ; Jakubowicz, J.</i>	
MHT: A LIGHT-WEIGHT SCALABLE ZERO-HOP MPI ENABLED DISTRIBUTED KEY-VALUE STORE.....	2901
<i>Xiaobing Zhou ; Tonglin Li ; Ke Wang ; Dongfang Zhao ; Sadooghi, I. ; Raicu, I.</i>	

BIG DATA: CLOUD COMPUTING IN GENOMICS APPLICATIONS	2904
<i>Hangu Yeo ; Crawford, C.H.</i>	
INTEGRATING SEMANTIC KNOWLEDGE INTO TAG-LDA MODEL THROUGH CLOUD MODEL	2907
<i>Maoyuan Zhang ; Fang Yuan ; Jianping Zhu</i>	
A CASE STUDY TO APPLY MOBILE TECHNOLOGY INTO INDIVIDUAL'S LOCAL COMMUNITY	2910
<i>Yunkai Liu ; Magno, C.</i>	
CLAIRVOYANT-PUSH: A REAL-TIME NEWS PERSONALIZED PUSH NOTIFIER USING TOPIC MODELING AND SOCIAL SCORING FOR ENHANCED READER ENGAGEMENT	2913
<i>Biyang Tan ; Sangaralingam, K. ; Singh, V.K. ; Saripaka, C.S. ; Manai, G.</i>	
USING PROBABILISTIC APPROACH TO JOINT CLUSTERING AND STATISTICAL INFERENCE: ANALYTICS FOR BIG INVESTMENT DATA	2916
<i>Hua Fang ; Honggang Wang ; Chonggang Wang ; Daneshmand, M.</i>	
TOWARDS A SUBGRAPH/SUPERGRAPH CACHED QUERY-GRAPH INDEX	2919
<i>Jing Wang ; Ntarmos, N. ; Triantafillou, P.</i>	
30 DAY HOSPITAL READMISSION ANALYSIS	2922
<i>Maddipatla, R.M. ; Hadzikadic, M. ; Misra, D.P. ; Lixia Yao</i>	
USING PAIRWISE DIFFERENCE FEATURES TO MEASURE TEMPORAL CHANGES IN THE MICROBIAL ECOLOGY	2925
<i>Yazdani, M. ; Smarr, L.</i>	
A TIMELINE VISUALIZATION SYSTEM FOR ROAD TRAFFIC BIG DATA	2928
<i>Imawan, A. ; Joonho Kwon</i>	
A NEW AREA TOURIST RANKING METHOD	2930
<i>Chareyron, G. ; Branchet, B. ; Jacquot, S.</i>	
TEXT RETRIEVAL BASED ON THE FEATURE CONVERSION OF VECTOR SPACE	2933
<i>Maoyuan Zhang ; Jianping Zhu ; Lijun Hua ; Fang Yuan</i>	
BIG DATA GATHERING AND MINING PIPELINES FOR CRM USING OPEN-SOURCE	2936
<i>Kang Li ; Deolalikar, V. ; Pradhan, N.</i>	
UNIFIED FRAMEWORK FOR CLINICAL DATA ANALYTICS (U-CDA)	2939
<i>Gholap, J. ; Janeja, V.P. ; Yesha, Y.</i>	
A NOVEL INITIALIZATION METHOD FOR PARTICLE SWARM OPTIMIZATION-BASED FCM IN BIG BIOMEDICAL DATA	2942
<i>Wang, C.J. ; Hua Fang ; Chonggang Wang ; Daneshmand, M. ; Honggang Wang</i>	
ALGORITHMIC CONTENT GENERATION FOR PRODUCTS	2945
<i>Khatri, C. ; Voleti, S. ; Veeraraghavan, S. ; Parikh, N. ; Islam, A. ; Mahmood, S. ; Garg, N. ; Singh, V.</i>	
HOTSPOTS OF NEWS ARTICLES: JOINT MINING OF NEWS TEXT & SOCIAL MEDIA TO DISCOVER CONTROVERSIAL POINTS IN NEWS	2948
<i>Lourentzou, I. ; Dyer, G. ; Sharma, A. ; ChengXiang Zhai</i>	
IMPROVING THE QUALITY OF SEMANTIC RELATIONSHIPS EXTRACTED FROM MASSIVE USER BEHAVIORAL DATA	2951
<i>AlJadda, K. ; Korayem, M. ; Grainger, T.</i>	
ANALYSIS OF STAR RATINGS IN CONSUMER REVIEWS: A CASE STUDY OF YELP	2954
<i>Prithivirajan, M. ; Lai, V. ; Kyong Jin Shim ; Koo Ping Shung</i>	
FROM STARS TO PATIENTS: LESSONS FROM SPACE SCIENCE AND ASTROPHYSICS FOR HEALTH CARE INFORMATICS	2957
<i>Djorgovski, S.G. ; Mahabal, A.A. ; Chaudhry, B. ; Crichton, D.J.</i>	
Author Index	