

2014 IEEE International Conference on Big Data

(Big Data 2014)

**Washington, DC, USA
27-30 October 2014**

Pages 1-682



IEEE Catalog Number: CFP14BGD-POD
ISBN: 978-1-4799-5667-8

TABLE OF CONTENTS

VOLUME 1

NEVER-ENDING LANGUAGE LEARNING	1
<i>Mitchell, T. ; Fredkin, E.</i>	
SMART DATA — HOW YOU AND I WILL EXPLOIT BIG DATA FOR PERSONALIZED DIGITAL HEALTH AND MANY OTHER ACTIVITIES	2
<i>Sheth, A.</i>	
ADDRESSING HUMAN BOTTLENECKS IN BIG DATA	4
<i>Hellerstein, J.M.</i>	
BASIC: AN ALTERNATIVE TO BASE FOR LARGE-SCALE DATA MANAGEMENT SYSTEM	5
<i>Lengdong Wu ; Li-Yan Yuan ; Jia-Huai You</i>	
BAYESWIPE: A MULTIMODAL SYSTEM FOR DATA CLEANING AND CONSISTENT QUERY ANSWERING ON STRUCTURED BIGDATA	15
<i>De, S. ; Yuheng Hu ; Yi Chen ; Kambhampati, S.</i>	
SCALING UP M-ESTIMATION VIA SAMPLING DESIGNS: THE HORVITZ-THOMPSON STOCHASTIC GRADIENT DESCENT	25
<i>Clemencon, S. ; Bertail, P. ; Chautru, E.</i>	
METADATA CAPITAL: SIMULATING THE PREDICTIVE VALUE OF SELF-GENERATED HEALTH INFORMATION (SGHI)	31
<i>Greenberg, J. ; Ogletree, A. ; Murillo, A.P. ; Caruso, T.P. ; Huang, H.</i>	
REPRESENTATIVE SUBSETS FOR BIG DATA LEARNING USING K-NN GRAPHS	37
<i>Mall, R. ; Jumutc, V. ; Langone, R. ; Suykens, J.A.K.</i>	
TOWARDS BUILDING AND EVALUATING A PERSONALIZED LOCATION-BASED RECOMMENDER SYSTEM	43
<i>Rubing Duan ; Goh, R.S.M. ; Feng Yang ; Yong Kiam Tan ; Valenzuela, J.F.B.</i>	
ON THE PERFORMANCE OF MAPREDUCE: A STOCHASTIC APPROACH	49
<i>Ahmed, S.T. ; Loguinov, D.</i>	
PGMHD: A SCALABLE PROBABILISTIC GRAPHICAL MODEL FOR MASSIVE HIERARCHICAL DATA PROBLEMS	55
<i>AlJadda, K. ; Korayem, M. ; Ortiz, C. ; Grainger, T. ; Miller, J.A. ; York, W.S.</i>	
FUSIONS: TOWARD SUPPORTING DATA-INTENSIVE SCIENTIFIC APPLICATIONS ON EXTREME-SCALE HIGH-PERFORMANCE COMPUTING SYSTEMS	61
<i>Dongfang Zhao ; Zhao Zhang ; Xiaobing Zhou ; Tonglin Li ; Ke Wang ; Kimpe, D. ; Carns, P. ; Ross, R. ; Raicu, I.</i>	
BURSTMEM: A HIGH-PERFORMANCE BURST BUFFER SYSTEM FOR SCIENTIFIC APPLICATIONS	71
<i>Teng Wang ; Oral, S. ; Yandong Wang ; Settlemyer, B. ; Atchley, S. ; Weikuan Yu</i>	
PARTIAL ROLLBACK-BASED SCHEDULING ON IN-MEMORY TRANSACTIONAL DATA GRIDS	80
<i>Junwhan Kim</i>	
DETECTING AND IDENTIFYING SYSTEM CHANGES IN THE CLOUD VIA DISCOVERY BY EXAMPLE	90
<i>Hao Chen ; Duri, S.S. ; Bala, V. ; Bila, N.T. ; Isci, C. ; Coskun, A.K.</i>	
PIGOUT: MAKING MULTIPLE HADOOP CLUSTERS WORK TOGETHER	100
<i>Kyungho Jeon ; Chandrashekhara, S. ; Feng Shen ; Mehra, S. ; Kennedy, O. ; Ko, S.Y.</i>	
PARALLEL BREADTH FIRST SEARCH ON GPU CLUSTERS	110
<i>Zhisong Fu ; Dasari, H.K. ; Bebee, B. ; Berzins, M. ; Thompson, B.</i>	
OPTIMIZING LOAD BALANCING AND DATA-LOCALITY WITH DATA-AWARE SCHEDULING	119
<i>Ke Wang ; Xraobing Zhou ; Tonglin Li ; Dongfang Zhao ; Lang, M. ; Raicu, I.</i>	
ONLINE TEMPORAL-SPATIAL ANALYSIS FOR DETECTION OF CRITICAL EVENTS IN CYBER-PHYSICAL SYSTEMS	129
<i>Zhang Fu ; Almgren, M. ; Landsiedel, O. ; Papatriantafilou, M.</i>	
A CROSS-JOB FRAMEWORK FOR MAPREDUCE SCHEDULING	135
<i>Xuejie Xiao ; Jian Tang ; Zhenhua Chen ; Jielong Xu ; Chonggang Wang</i>	
SCHEDULING MAPREDUCE TASKS ON VIRTUAL MAPREDUCE CLUSTERS FROM A TENANT'S PERSPECTIVE	141
<i>Jia-Chun Lin ; Ming-Chang Lee ; Yahyapour, R.</i>	

FLEXDAS: A FLEXIBLE DIRECT ATTACHED STORAGE FOR I/O INTENSIVE APPLICATIONS	147
<i>Ono, T. ; Konishi, Y. ; Tanimoto, T. ; Iwamatsu, N. ; Miyoshi, T. ; Tanaka, J.</i>	
A TWO-SIDED MARKET MECHANISM FOR TRADING BIG DATA COMPUTING COMMODITIES	153
<i>Mashayekhy, L. ; Nejad, M.M. ; Grosu, D.</i>	
MMAP: FAST BILLION-SCALE GRAPH COMPUTATION ON A PC VIA MEMORY MAPPING	159
<i>Zhiyuan Lin ; Minsuk Kahng ; Sabrin, K.M. ; Chau, D.H.P. ; Ho Lee ; Kang, U.</i>	
LARGE-SCALE NETWORK TRAFFIC MONITORING WITH DBSTREAM, A SYSTEM FOR ROLLING BIG DATA ANALYSIS	165
<i>Bar, A. ; Finamore, A. ; Casas, P. ; Golab, L. ; Mellia, M.</i>	
SYNTHETIC DATA GENERATION FOR THE INTERNET OF THINGS	171
<i>Anderson, J.W. ; Kennedy, K.E. ; Ngo, L.B. ; Luckow, A. ; Apon, A.W.</i>	
EVALUATING THE PERFORMANCE AND SCALABILITY OF THE CEPH DISTRIBUTED STORAGE SYSTEM	177
<i>Gudu, D. ; Hardt, M. ; Streit, A.</i>	
INCREMENTAL WINDOW AGGREGATES OVER ARRAY DATABASE	183
<i>Li Jiang ; Kawashima, H. ; Tatebe, O.</i>	
BIGCACHE FOR BIG-DATA SYSTEMS	189
<i>Roger, M.A. ; Yiqi Xu ; Ming Zhao</i>	
AUTOMATED WORKLOAD-AWARE ELASTICITY OF NOSQL CLUSTERS IN THE CLOUD	195
<i>Kassela, E. ; Boumpouka, C. ; Konstantinou, I. ; Koziris, N.</i>	
DISTRIBUTED CLASS DEPENDENT FEATURE ANALYSIS — A BIG DATA APPROACH	201
<i>Khoa Luu ; Chenchen Zhu ; Savvides, M.</i>	
VENU: ORCHESTRATING SSDS IN HADOOP STORAGE	207
<i>Krish, K.R. ; Iqbal, M.S. ; Butt, A.R.</i>	
IN-MEMORY I/O AND REPLICATION FOR HDFS WITH MEMCACHED: EARLY EXPERIENCES	213
<i>Islam, N.S. ; Xiaoyi Lu ; Wasi-ur-Rahman, M. ; Rajachandrasekar, R. ; Panda, D.K.D.K.</i>	
ENABLING COMPOSITE APPLICATIONS THROUGH AN ASYNCHRONOUS SHARED MEMORY INTERFACE	219
<i>Orstott, D. ; Evans, N. ; Ionkov, L. ; Ming Zhao ; Lang, M.</i>	
K-BALANCED SORTING AND SKEW JOIN IN MPI AND MAPREDUCE	225
<i>Silu Huang ; Fu, W.-C.</i>	
VIRTUAL CHUNKS: ON SUPPORTING RANDOM ACCESSES TO SCIENTIFIC DATA IN COMPRESSIBLE STORAGE SYSTEMS	231
<i>Dongfang Zhao ; Jian Yin ; Kan Qiao ; Raicu, I.</i>	
EXAMINATION OF DATA, RULE GENERATION AND DETECTION OF PHISHING URLS USING ONLINE LOGISTIC REGRESSION	241
<i>Feroz, M.N. ; Mengel, S.</i>	
MAIN MEMORY EVALUATION OF RECURSIVE QUERIES ON MULTICORE MACHINES	251
<i>Mohan Yang ; Zaniolo, C.</i>	
PREDICTING GLAUCOMA PROGRESSION USING MULTI-TASK LEARNING WITH HETEROGENEOUS FEATURES	261
<i>Maya, S. ; Morino, K. ; Yamanishi, K.</i>	
PROVENANCE-BASED OBJECT STORAGE PREDICTION SCHEME FOR SCIENTIFIC BIG DATA APPLICATIONS	271
<i>Dong Dai ; Yong Chen ; Kimpe, D. ; Ross, R.</i>	
SYNERGISTIC PARTITIONING IN MULTIPLE LARGE SCALE SOCIAL NETWORKS	281
<i>Songchang Jin ; Jiawei Zhang ; Yu, P.S. ; Shuqiang Yang ; Aiping Li</i>	
TRISTAN: REAL-TIME ANALYTICS ON MASSIVE TIME SERIES USING SPARSE DICTIONARY COMPRESSION	291
<i>Marascu, A. ; Pompey, P. ; Bouillet, E. ; Wurst, M. ; Verscheure, O. ; Grund, M. ; Cudre-Mauroux, P.</i>	
PERFORMANCE MODELING IN CUDA STREAMS — A MEANS FOR HIGH-THROUGHPUT DATA PROCESSING	301
<i>Hao Li ; Di Yu ; Kumar, A. ; Yi-Cheng Tu</i>	
MINIMIZING DATA MOVEMENT THROUGH QUERY TRANSFORMATION	311
<i>Leyshock, P. ; Maier, D. ; Tufte, K.</i>	
MULTILEVEL PARTITIONING OF LARGE UNSTRUCTURED GRIDS	317
<i>Akande, O.O. ; Rhodes, P.J.</i>	
LOW COMPLEXITY SENSING FOR BIG SPATIO-TEMPORAL DATA	323
<i>Donggeun Lee ; Jaesik Choi</i>	

IN-ADVANCE DATA ANALYTICS FOR REDUCING TIME TO DISCOVERY	329
<i>Jialin Liu ; Yin Lu ; Yong Chen</i>	
ESTIMATING PAIRWISE DISTANCES IN LARGE GRAPHS.....	335
<i>Christoforaki, M. ; Suel, T.</i>	
DISTRIBUTED ADAPTIVE MODEL RULES FOR MINING BIG DATA STREAMS.....	345
<i>Anh Thu Vu ; De Francisci Morales, G. ; Gama, J. ; Bifet, A.</i>	
SPARSE COMPUTATION FOR LARGE-SCALE DATA MINING.....	354
<i>Hochbaum, D.S. ; Baumann, P.</i>	
TOPIC SIMILARITY NETWORKS: VISUAL ANALYTICS FOR LARGE DOCUMENT SETS.....	364
<i>Maiya, A.S. ; Rolfe, R.M.</i>	
EFFICIENT BREADTH-FIRST SEARCH ON A HETEROGENEOUS PROCESSOR	373
<i>Daga, M. ; Nutter, M. ; Meswani, M.</i>	
WEB-BASED VISUAL ANALYTICS FOR EXTREME SCALE CLIMATE SCIENCE	383
<i>Steed, C.A. ; Evans, K.J. ; Harney, J.F. ; Jewell, B.C. ; Shipman, G. ; Smith, B.E. ; Thornton, P.E. ; Williams, D.N.</i>	
GEOTAGGING ONE HUNDRED MILLION TWITTER ACCOUNTS WITH TOTAL VARIATION MINIMIZATION	393
<i>Compton, R. ; Jurgens, D. ; Allen, D.</i>	
MEETING PREDICTABLE BUFFER LIMITS IN THE PARALLEL EXECUTION OF EVENT PROCESSING OPERATORS.....	402
<i>Mayer, R. ; Koldehofe, B. ; Rothermel, K.</i>	
METADATA EXTRACTION AND CORRECTION FOR LARGE-SCALE TRAFFIC SURVEILLANCE VIDEOS	412
<i>Xiaomeng Zhao ; Huadong Ma ; Haitao Zhang ; Yi Tang ; Guangping Fu</i>	
FACILITATING TWITTER DATA ANALYTICS: PLATFORM, LANGUAGE AND FUNCTIONALITY	421
<i>Ke Tao ; Hauff, C. ; Houben, G.-J. ; Abel, F. ; Wachsmuth, G.</i>	
VISUAL FUSION OF MEGA-CITY BIG DATA: AN APPLICATION TO TRAFFIC AND TWEETS DATA ANALYSIS OF METRO PASSENGERS	431
<i>Itoh, M. ; Yokoyama, D. ; Toyoda, M. ; Tomita, Y. ; Kawamura, S. ; Kitsuregawa, M.</i>	
GRAPHIQL: A GRAPH INTUITIVE QUERY LANGUAGE FOR RELATIONAL DATABASES	441
<i>Jindal, A. ; Madden, S.</i>	
EVALUATING DENSITY-BASED MOTION FOR BIG DATA VISUAL ANALYTICS	451
<i>Etemadpour, R. ; Murray, P. ; Forbes, A.G.</i>	
REGRESSION TREES FOR STREAMING DATA WITH LOCAL PERFORMANCE GUARANTEES	461
<i>Johansson, U. ; Sonstrod, C. ; Linusson, H. ; Bostrom, H.</i>	
DISTRIBUTED ALGORITHMS FOR K-TRUSS DECOMPOSITION	471
<i>Pei-Ling Chen ; Chung-Kuang Chou ; Ming-Syan Chen</i>	
PULP: SCALABLE MULTI-OBJECTIVE MULTI-CONSTRAINT PARTITIONING FOR SMALL-WORLD NETWORKS	481
<i>Slota, G.M. ; Madduri, K. ; Rajamanickam, S.</i>	
EFFECTIVE CACHING TECHNIQUES FOR ACCELERATING PATTERN MATCHING QUERIES	491
<i>Fard, A. ; Manda, S. ; Ramaswamy, L. ; Miller, J.A.</i>	
CLIQUE GUIDED COMMUNITY DETECTION.....	500
<i>Palsetia, D. ; Patwary, M.M.A. ; Hendrix, W. ; Agrawal, A. ; Choudhary, A.</i>	
LARGE-SCALE DISTRIBUTED SORTING FOR GPU-BASED HETEROGENEOUS SUPERCOMPUTERS.....	510
<i>Shamoto, H. ; Shirahata, K. ; Drozd, A. ; Sato, H. ; Matsuoka, S.</i>	
LARGE-SCALE LOGISTIC REGRESSION AND LINEAR SUPPORT VECTOR MACHINES USING SPARK.....	519
<i>Chieh-Yen Lin ; Cheng-Hao Tsai ; Ching-Pei Lee ; Chih-Jen Lin</i>	
NVM-BASED HYBRID BFS WITH MEMORY EFFICIENT DATA STRUCTURE	529
<i>Iwabuchi, K. ; Sato, H. ; Yasui, Y. ; Fujisawa, K. ; Matsuoka, S.</i>	
IDENTIFICATION OF SNP INTERACTIONS USING DATA-PARALLEL PRIMITIVES ON GPUS	539
<i>Altinigneli, C. ; Konten, B. ; Rujescir, D. ; Bohm, C. ; Plant, C.</i>	
RANDOM WALKS ON ADJACENCY GRAPHS FOR MINING LEXICAL RELATIONS FROM BIG TEXT DATA.....	549
<i>Shan Jiang ; ChengXiang Zhai</i>	
ENTITY RESOLUTION USING INFERRED RELATIONSHIPS AND BEHAVIOR.....	555
<i>Mugan, J. ; Chari, R. ; Hitt, L. ; McDermid, E. ; Sowell, M. ; Yuan Qu ; Coffman, T.</i>	

RAINBOW: A DISTRIBUTED AND HIERARCHICAL RDF TRIPLE STORE WITH DYNAMIC SCALABILITY	561
<i>Rong Gu ; Wei Hu ; Yihua Huang</i>	
IN-SITU VISUALIZATION AND COMPUTATIONAL STEERING FOR LARGE-SCALE SIMULATION OF TURBULENT FLOWS IN COMPLEX GEOMETRIES	567
<i>Hong Yi ; Rasquin, M. ; Jun Fang ; Bolotnov, I.A.</i>	
BUILDING K-NN GRAPHS FROM LARGE TEXT DATA	573
<i>Debatty, T. ; Michiardi, P. ; Thonnard, O. ; Mees, W.</i>	
LEARNING TO PREDICT SUBJECT-LINE OPENS FOR LARGE-SCALE EMAIL MARKETING	579
<i>Balakrishnan, R. ; Parekh, R.</i>	
MAGE: MATCHING APPROXIMATE PATTERNS IN RICHLY-ATTRIBUTED GRAPHS	585
<i>Pienta, R. ; Tamersoy, A. ; Hanghang Tong ; Duen Horng Chau</i>	
BOOTSTRAPPING K-MEANS FOR BIG DATA ANALYSIS	591
<i>Jungkyu Han ; Min Luo</i>	
DISTRIBUTED ADAPTIVE IMPORTANCE SAMPLING ON GRAPHICAL MODELS USING MAPREDUCE	597
<i>Haque, A. ; Chandra, S. ; Khan, L. ; Aggarwal, C.</i>	
KNOWLEDGE-BASED CLUSTERING OF SHIP TRAJECTORIES USING DENSITY-BASED APPROACH	603
<i>Bo Liu ; de Souza, E.N. ; Matwin, S. ; Sydow, M.</i>	
IMMERSIVE AND COLLABORATIVE DATA VISUALIZATION USING VIRTUAL REALITY PLATFORMS	609
<i>Donalek, C. ; Djorgovski, S.G. ; Cioc, A. ; Wang, A. ; Zhang, J. ; Lawler, E. ; Yeh, S. ; Mahabal, A. ; Graham, M. ; Drake, A. ; Davidoff, S. ; Norris, J.S. ; Longo, G.</i>	
THE ADAPTIVE PROJECTION FOREST: USING ADJUSTABLE EXCLUSION AND PARALLELISM IN METRIC SPACE INDEXES	615
<i>Thompson, L.P. ; Weijia Xu ; Miranker, D.P.</i>	
SCALING UP PRIORITIZED GRAMMAR ENUMERATION FOR SCIENTIFIC DISCOVERY IN THE CLOUD	621
<i>Worm, T. ; Chiu, K.</i>	
MR-TRIAGE: SCALABLE MULTI-CRITERIA CLUSTERING FOR BIG DATA SECURITY INTELLIGENCE APPLICATIONS	627
<i>Yun Shen ; Thonnard, O.</i>	
INCREASING THE VERACITY OF EVENT DETECTION ON SOCIAL MEDIA NETWORKS THROUGH USER TRUST MODELING	636
<i>Bodnar, T. ; Tucker, C. ; Hopkinson, K. ; Bilen, S.G.</i>	
EMPOWERING USERS OF SOCIAL NETWORKS TO ASSESS THEIR PRIVACY RISKS	644
<i>Estivill-Castro, V. ; Hough, P. ; Islam, M.Z.</i>	
A UNIFIED APPROACH TO NETWORK ANOMALY DETECTION	650
<i>Babaie, T. ; Chawla, S. ; Ardon, S. ; Yue Yu</i>	
E-SKETCH: GATHERING LARGE-SCALE ENERGY CONSUMPTION DATA BASED ON CONSUMPTION PATTERNS	656
<i>Zhichuan Huang ; Hongyao Luo ; Skoda, D. ; Ting Zhu ; Yu Gu</i>	
HIERARCHICAL MANAGEMENT OF LARGE-SCALE MALWARE DATA	666
<i>Kellogg, L. ; Ruttenberg, B. ; O'Connor, A. ; Howard, M. ; Pfeffer, A.</i>	
RANDOM PROJECTION BASED CLUSTERING FOR POPULATION GENOMICS	675
<i>Tasoulis, S. ; Lu Cheng ; Valimaki, N. ; Croucher, N.J. ; Harris, S.R. ; Hanage, W.P. ; Roos, T. ; Corander, J.</i>	

VOLUME 2

STRUCTURE RECOGNITION FROM HIGH RESOLUTION IMAGES OF CERAMIC COMPOSITES	683
<i>Ushizima, D. ; Perciano, T. ; Krishnan, H. ; Loring, B. ; Bale, H. ; Parkinson, D. ; Sethian, J.</i>	
COMBINING HADOOP AND GPU TO PREPROCESS LARGE AFFYMETRIX MICROARRAY DATA	692
<i>Sufeng Niu ; Guangyu Yang ; Sarma, N. ; Pengfei Xuan ; Smith, M.C. ; Srimani, P. ; Feng Luo</i>	
CONTENT-BASED ACCESS CONTROL: USE DATA CONTENT TO ASSIST ACCESS CONTROL FOR LARGE-SCALE CONTENT-CENTRIC DATABASES	701
<i>Wenrong Zeng ; Yuhao Yang ; Bo Luo</i>	
LOCATING VISUAL STORM SIGNATURES FROM SATELLITE IMAGES	711
<i>Yu Zhang ; Wistar, S. ; Piedra-Fernandez, J.A. ; Jia Li ; Steinberg, M.A. ; Wang, J.Z.</i>	

ACCURATE AND EFFICIENT SELECTION OF THE BEST CONSUMPTION PREDICTION METHOD IN SMART GRIDS	721
<i>Frincu, M. ; Chelmis, C. ; Noor, M.U. ; Prasanna, V.</i>	
VISUALIZATIONS FOR SENSE-MAKING IN FINANCIAL MARKET REGULATION	730
<i>Todd, A. ; Scherer, W. ; Beling, P. ; Paddrik, M. ; Haynes, R.</i>	
BIG AUTOMOTIVE DATA: LEVERAGING LARGE VOLUMES OF DATA FOR KNOWLEDGE-DRIVEN PRODUCT DEVELOPMENT	736
<i>Johanson, M. ; Belenki, S. ; Jalminger, J. ; Fant, M. ; Gjertz, M.</i>	
ON THE IMPACT OF SOCIO-ECONOMIC FACTORS ON POWER LOAD FORECASTING.....	742
<i>Han, Y. ; Sha, X. ; Grover-Silva, E. ; Michiardi, P.</i>	
TOWARD PERSONALIZED AND SCALABLE VOICE-ENABLED SERVICES POWERED BY BIG DATA	748
<i>Jong Hoon Ahnn</i>	
MAPLE: A MAPREDUCE PIPELINE FOR LATTICE-BASED EVALUATION AND ITS APPLICATION TO SNOMED CT	754
<i>Guo-Qiang Zhang ; Wei Zhu ; Mengmeng Sun ; Shiqiang Tao ; Bodenreider, O. ; Licong Cui</i>	
DYNAMIC PRE-TRAINING OF DEEP RECURRENT NEURAL NETWORKS FOR PREDICTING ENVIRONMENTAL MONITORING DATA.....	760
<i>Ong, B.T. ; Sugiura, K. ; Zettsu, K.</i>	
PERLDOOP: EFFICIENT EXECUTION OF PERL SCRIPTS ON HADOOP CLUSTERS.....	766
<i>Abuin, J.M. ; Pichel, J.C. ; Pena, T.F. ; Gamallo, P. ; Garci?a, M.</i>	
DEPARTMENT OF ENERGY STRATEGIC ROADMAP FOR EARTH SYSTEM SCIENCE DATA INTEGRATION	772
<i>Williams, D.N. ; Palanisamy, G. ; Shipman, G. ; Boden, T.A. ; Voyles, J.W.</i>	
ANALYZING THE LANGUAGE OF FOOD ON SOCIAL MEDIA	778
<i>Fried, D. ; Surdeanu, M. ; Kobourov, S. ; Hingle, M. ; Bell, D.</i>	
USING GEOMETRIC STRUCTURES TO IMPROVE THE ERROR CORRECTION ALGORITHM OF HIGH-THROUGHPUT SEQUENCING DATA ON MAPREDUCE FRAMEWORK	784
<i>Wei-Chun Chung ; Yu-Jung Chang ; Lee, D.T. ; Jan-Ming Ho</i>	
EMPOWERING PERSONALIZED MEDICINE WITH BIG DATA AND SEMANTIC WEB TECHNOLOGY: PROMISES, CHALLENGES, AND USE CASES	790
<i>Panahiazar, M. ; Taslimitehrani, V. ; Jadhav, A. ; Pathak, J.</i>	
ON SCALING TIME DEPENDENT SHORTEST PATH COMPUTATIONS FOR DYNAMIC TRAFFIC ASSIGNMENT	796
<i>Gupta, A. ; Weijia Xu ; Perrine, K. ; Bell, D. ; Ruiz-Juri, N.</i>	
HIGH VOLUME GEOSPATIAL MAPPING FOR INTERNET-OF-VEHICLE SOLUTIONS WITH IN-MEMORY MAP-REDUCE PROCESSING.....	802
<i>Tao Zhong ; Doshi, K. ; Gang Deng ; Xiaoming Yang ; Hegao Zhang</i>	
CROWDSOURCED QUERY AUGMENTATION THROUGH SEMANTIC DISCOVERY OF DOMAIN-SPECIFIC JARGON.....	808
<i>AlJadda, K. ; Korayem, M. ; Grainger, T. ; Russell, C.</i>	
SPATIAL COMPUTATIONS OVER TERABYTE-SIZED IMAGES ON HADOOP PLATFORMS	816
<i>Bajcsy, P. ; Nguyen, P. ; Vandecreme, A. ; Brady, M.</i>	
RECALL ESTIMATION FOR RARE TOPIC RETRIEVAL FROM LARGE CORPUSES	825
<i>Bommanavar, P. ; Kolcz, A. ; Rajaraman, A.</i>	
LIGHTWEIGHT APPROXIMATE TOP-K FOR DISTRIBUTED SETTINGS	835
<i>Deolalikar, V. ; Eshghi, K.</i>	
QUERY REVISION DURING CLUSTER BASED SEARCH ON LARGE UNSTRUCTURED CORPORA.....	845
<i>Deolalikar, V.</i>	
ASTRO: A PREDICTIVE MODEL FOR ANOMALY DETECTION AND FEEDBACK-BASED SCHEDULING ON HADOOP	854
<i>Gupta, C. ; Bansal, M. ; Tzu-Cheng Chuang ; Sinha, R. ; Ben-romdhane, S.</i>	
AUTOMATING DATA INTEGRATION WITH HIPERFUSE	863
<i>Huang, E. ; Quiroz, A. ; Ceriani, L.</i>	
RECOMMENDING SIMILAR ITEMS IN LARGE-SCALE ONLINE MARKETPLACES	868
<i>Katukuri, J. ; Konik, T. ; Mukherjee, R. ; Kolay, S.</i>	
SE-CDA: A SCALABLE AND EFFICIENT COMMUNITY DETECTION ALGORITHM	877
<i>Lunagariya, D.C. ; Somayajulu, D.V.L.N. ; Krishna, P.R.</i>	

INCREASING THE ACCESSIBILITY TO BIG DATA SYSTEMS VIA A COMMON SERVICES API	883
<i>Malcolm, R. ; Morrison, C. ; Grandison, T. ; Thorpe, S. ; Christie, K. ; Wallace, A. ; Green, D. ; Jarrett, J. ; Campbell, A.</i>	
BIG DATA PREDICTIVE ANALYTICS FOR PROACTIVE SEMICONDUCTOR EQUIPMENT MAINTENANCE.....	893
<i>Munirathinam, S. ; Ramadoss, B.</i>	
FUTURE DIRECTIONS OF HUMANS IN BIG DATA RESEARCH: SUMMARY OF THE 1ST WORKSHOP ON HUMAN-CENTERED BIG DATA RESEARCH	903
<i>Paul, Celeste Lyn ; Argenta, Chris ; Elm, William ; Endert, Alex</i>	
ALOJA: A SYSTEMATIC STUDY OF HADOOP DEPLOYMENT VARIABLES TO ENABLE AUTOMATED CHARACTERIZATION OF COST-EFFECTIVENESS	905
<i>Poggi, N. ; Carrera, D. ; Call, A. ; Mendoza, S. ; Becerra, Y. ; Torres, J. ; Ayguade, E. ; Gagliardi, F. ; Labarta, J. ; Reinauer, R. ; Vujic, N. ; Green, D. ; Blakeley, J.</i>	
HETEROGENEOUS STREAM PROCESSING FOR DISASTER DETECTION AND ALARMING	914
<i>Schnizler, F. ; Liebig, T. ; Marmor, S. ; Souto, G. ; Bothe, S. ; Stange, H.</i>	
IDENTIFYING TOP CHINESE NETWORK BUZZWORDS FROM SOCIAL MEDIA BIG DATA SET BASED ON TIME-DISTRIBUTION FEATURES	924
<i>Yongli Tang ; Tingting He ; Bo Li ; Xiaohua Hu</i>	
BRIDGING HIGH VELOCITY AND HIGH VOLUME INDUSTRIAL BIG DATA THROUGH DISTRIBUTED IN-MEMORY STORAGE & ANALYTICS	932
<i>Williams, J.W. ; Aggour, K.S. ; Interrante, J. ; McHugh, J. ; Pool, E.</i>	
GRAPH ANALYTICS AND STORAGE.....	942
<i>Yinglong Xia ; Tanase, I.G. ; Lifeng Nai ; Wei Tan ; Yanbin Liu ; Crawford, J. ; Ching-Yung Lin</i>	
AN INITIAL STUDY OF PREDICTIVE MACHINE LEARNING ANALYTICS ON LARGE VOLUMES OF HISTORICAL DATA FOR POWER SYSTEM APPLICATIONS.....	952
<i>Jiang Zheng ; Dagnino, A.</i>	
IN UNITY THERE IS STRENGTH: SHOWCASING A UNIFIED BIG DATA PLATFORM WITH MAPREDUCE OVER BOTH OBJECT AND FILE STORAGE.....	960
<i>Rui Zhang ; Hildebrand, D. ; Tewari, R.</i>	
TOWARD SMART MANUFACTURING USING DECISION ANALYTICS.....	967
<i>Brodsky, A. ; Krishnamoorthy, M. ; Menasce, D.A. ; Guodong Shao ; Rachuri, S.</i>	
AN INTELLIGENT MACHINE MONITORING SYSTEM FOR ENERGY PREDICTION USING A GAUSSIAN PROCESS REGRESSION	978
<i>Bhinge, R. ; Biswas, N. ; Dornfeld, D. ; Jinkyoo Park ; Law, K.H. ; Helu, M. ; Rachuri, S.</i>	
TOWARDS A DOMAIN-SPECIFIC FRAMEWORK FOR PREDICTIVE ANALYTICS IN MANUFACTURING.....	987
<i>Lechevalier, D. ; Narayanan, A. ; Rachuri, S.</i>	
UNCERTAINTY QUANTIFICATION IN PERFORMANCE EVALUATION OF MANUFACTURING PROCESSES.....	996
<i>Nannapaneni, S. ; Mahadevan, S.</i>	
CLOUDMAN: A PLATFORM FOR PORTABLE CLOUD MANUFACTURING SERVICES	1006
<i>Qanbari, S. ; Zadeh, S.M. ; Vedaiei, S. ; Dustdar, S.</i>	
BUILDING A RIGOROUS FOUNDATION FOR PERFORMANCE ASSURANCE ASSESSMENT TECHNIQUES FOR “SMART” MANUFACTURING SYSTEMS	1015
<i>Roy, U. ; Yunpeng Li ; Bicheng Zhu</i>	
A SYSTEM ARCHITECTURE FOR MANUFACTURING PROCESS ANALYSIS BASED ON BIG DATA AND PROCESS MINING TECHNIQUES	1024
<i>Hanna Yang ; Minjeong Park ; Minsu Cho ; Minseok Song ; Seongjoo Kim</i>	
RESEARCHING PERSONS & ORGANIZATIONS: AWAKE: FROM TEXT TO AN ENTITY-CENTRIC KNOWLEDGE BASE.....	1030
<i>Boschee, E. ; Freedman, M. ; Khanwalkar, S. ; Kumar, A. ; Srivastava, A. ; Weischedel, R.</i>	
INTEGRATING EXISTING LARGE SCALE MEDICAL LABORATORY DATA INTO THE SEMANTIC WEB FRAMEWORK	1040
<i>Al Haider, N. ; Abidi, S. ; van Woensel, W. ; Abidi, S.S.R.</i>	
PATH KNOWLEDGE DISCOVERY: ASSOCIATION MINING BASED ON MULTI-CATEGORY LEXICONS	1049
<i>Chen Liu ; Chu, W.W. ; Sabb, F. ; Parker, D.S. ; Korpela, J.</i>	
STOCHASTIC FINITE AUTOMATA FOR THE TRANSLATION OF DNA TO PROTEIN.....	1060
<i>Tsau-Young Lin ; Shah, A.H.</i>	

DATA MINING AND SHARING TOOL FOR HIGH CONTENT SCREENING LARGE SCALE BIOLOGICAL IMAGE DATA	1068
<i>Shah, A.H. ; Gopalakrishnan, G. ; Rajendran, A. ; Liebel, U.</i>	
A BUILDING PERFORMANCE EVALUATION & VISUALIZATION SYSTEM.....	1077
<i>Stavropoulos, G. ; Krinidis, S. ; Ioannidis, D. ; Moustakas, K. ; Tzovaras, D.</i>	
STATISTICAL TECHNIQUE FOR ONLINE ANOMALY DETECTION USING SPARK OVER HETEROGENEOUS DATA FROM MULTI-SOURCE VMWARE PERFORMANCE DATA	1086
<i>Solaimani, M. ; Iftekhara, M. ; Khan, L. ; Thuraisingham, B.</i>	
EXTRACTING DISCRIMINATIVE SHAPELETS FROM HETEROGENEOUS SENSOR DATA	1095
<i>Patri, O.P. ; Sharma, A.B. ; Haifeng Chen ; Guofei Jiang ; Panangadan, A.V. ; Prasanna, V.K.</i>	
WHY NAME AMBIGUITY RESOLUTION MATTERS FOR SCHOLARLY BIG DATA RESEARCH	1105
<i>Jinseok Kim ; Diesner, J. ; Aleyasen, A. ; Heejun Kim ; Hwan-Min Kim</i>	
EVOLUTION OF SCIENTIFIC COLLABORATION NETWORKS	1111
<i>Madaan, G. ; Jolad, S.</i>	
THE OCEANLINK PROJECT	1118
<i>Narock, T. ; Krisnadhi, A. ; Hitzler, P. ; Cheatham, M. ; Arko, R. ; Carbotte, S. ; Shepherd, A. ; Chandler, C. ; Raymond, L. ; Wiebe, P. ; Finin, T.</i>	
MANAGING THE ACADEMIC DATA LIFECYCLE: A CASE STUDY OF HPCC.....	1126
<i>Payne, M.E. ; Ngo, L.B. ; Villanustre, F. ; Apon, A.W.</i>	
SCALABLE BIG DATA COMPUTING FOR THE PERSONALIZATION OF MACHINE LEARNED MODELS AND ITS APPLICATION TO AUTOMATIC SPEECH RECOGNITION SERVICE	1135
<i>Ahnn, J.H.</i>	
COMPUTING FUZZY ROUGH APPROXIMATIONS IN LARGE SCALE INFORMATION SYSTEMS	1143
<i>Asfoor, H. ; Srinivasan, R. ; Vasudevan, G. ; Verbiest, N. ; Cornells, C. ; Tolentino, M. ; Teredesai, A. ; De Cock, M.</i>	
FAST LEARNING FOR BIG DATA APPLICATIONS USING PARAMETERIZED MULTILAYER PERCEPTRON.....	1151
<i>Chandra, B. ; Sharma, R.K.</i>	
CALCULATING FEATURE IMPORTANCE IN DATA STREAMS WITH CONCEPT DRIFT USING ONLINE RANDOM FOREST	1157
<i>Cassidy, A.P. ; Deviney, F.A.</i>	
TOWARDS SCALABLE GRAPH COMPUTATION ON MOBILE DEVICES.....	1163
<i>Yiqi Chen ; Zhiyuan Lin ; Pienta, R. ; Minsuk Kahng ; Duen Horng Chau</i>	
BOOSTING STOCHASTIC NEWTON DESCENT FOR BIGDATA LARGE SCALE CLASSIFICATION.....	1170
<i>D'Ambrosio, R. ; Belhajali, W. ; Barlaud, M.</i>	
FEATURE SELECTION FOR TEXT CLUSTERING IN LIMITED MEMORY USING MONTE CARLO WRAPPER.....	1176
<i>Deolalikar, V.</i>	
WS2F: A WEAKLY SUPERVISED FRAMEWORK FOR DATA STREAM FILTERING	1184
<i>Cailing Dong ; Agarwal, A.</i>	
AN IMPROVED MEMORY MANAGEMENT SCHEME FOR LARGE SCALE GRAPH COMPUTING ENGINE GRAPHCHI.....	1192
<i>Yifang Jiang ; Diao Zhang ; Kai Chen ; Qu Zhou ; Yi Zhou ; Jianhua He</i>	
FAST ALGORITHM FOR COMPUTING WEIGHTED PROJECTION QUANTILES AND DATA DEPTH FOR HIGH-DIMENSIONAL LARGE DATA CLOUDS.....	1198
<i>Mukherjee, U.K. ; Chatterjee, S.</i>	
FS3: A SAMPLING BASED METHOD FOR TOP-K FREQUENT SUBGRAPH MINING.....	1206
<i>Saha, T.K. ; Al Hasan, M.</i>	
A CLUSTERING BASED SCALABLE HYBRID APPROACH FOR WEB PAGE RECOMMENDATION.....	1214
<i>Sharif, M.A. ; Raghavan, V.V.</i>	
MULTIRESOLUTION ANALYSIS OF INCOMPLETE RANKINGS WITH APPLICATIONS TO PREDICTION.....	1222
<i>Sibony, E. ; Clemencon, S. ; Jakubowicz, J.</i>	
PAIRWISE TOPIC MODEL VIA RELATION EXTRACTION	1230
<i>Xiaoli Song ; Yue Shang ; Yuan Ling ; Mengwen Liu ; Xiaohua Hu</i>	

A MULTI-VIEW TWO-LEVEL CLASSIFICATION METHOD FOR GENERALIZED MULTI-INSTANCE PROBLEMS	1238
<i>Xiaoguang Wang ; Xuan Liu ; Matwin, S. ; Japkowicz, N. ; Hongyu Guo</i>	
APPLYING INSTANCE-WEIGHTED SUPPORT VECTOR MACHINES TO CLASS IMBALANCED DATASETS.....	1246
<i>Xiaoguang Wang ; Xuan Liu ; Matwin, S. ; Japkowicz, N.</i>	
CONNECTING THE DOTS: TRIANGLE COMPLETION AND RELATED PROBLEMS ON LARGE DATA SETS USING GPUS	1253
<i>Chatterjee, A. ; Radhakrishnan, S. ; Sekharan, C.N.</i>	
PARK: AN EFFICIENT ALGORITHM FOR K-CORE DECOMPOSITION ON MULTICORE PROCESSORS.....	1261
<i>Dasari, N.S. ; Desh, R. ; Zubair, M.</i>	
A PARTITIONING APPROACH TO SCALING ANOMALY DETECTION IN GRAPH STREAMS	1269
<i>Eberle, W. ; Holder, L.</i>	
FRACTIONAL GREEDY AND PARTIAL RESTREAMING PARTITIONING: NEW METHODS FOR MASSIVE GRAPH PARTITIONING	1277
<i>Echbarthi, G. ; Kheddouci, H.</i>	
GLOBAL GRAPHS: A MIDDLEWARE FOR LARGE SCALE GRAPH PROCESSING	1285
<i>Faisal, S.M. ; Parthasarathy, S. ; Sadayappan, P.</i>	
TOWARD AN EFFICIENT, HIGHLY SCALABLE MAXIMUM CLIQUE SOLVER FOR MASSIVE GRAPHS	1293
<i>Hagan, R.D. ; Phillips, C.A. ; Wang, K. ; Rogers, G.L. ; Langsten, M.A.</i>	
EXTENDING SPARQL WITH GRAPH FUNCTIONS	1298
<i>Mizell, D. ; Maschhoff, K.J. ; Reinhardt, S.P.</i>	
CHANGE DETECTION IN TEMPORALLY EVOLVING COMPUTER NETWORKS: A BIG DATA FRAMEWORK	1306
<i>Namayanja, J.M. ; Janeja, V.P.</i>	
DETECTING COMMUNITIES AROUND SEED NODES IN COMPLEX NETWORKS.....	1314
<i>Staudt, C.L. ; Marrakchi, Y. ; Meyerhenke, H.</i>	
ACCESS-AVERSE FRAMEWORK FOR COMPUTING LOW-RANK MATRIX APPROXIMATIONS.....	1322
<i>Yamazaki, I. ; Mary, T. ; Kurzak, J. ; Tomov, S. ; Dongarra, J.</i>	
ARCHITECTURE-AWARE GRAPH REPARTITIONING FOR DATA-INTENSIVE SCIENTIFIC COMPUTING	1330
<i>Angen Zheng ; Labrinidis, A. ; Chrysanthis, P.K.</i>	
A NEW ZIGZAG MDS CODE WITH OPTIMAL ENCODING AND EFFICIENT DECODING	1338
<i>Jun Chen ; Hui Li ; Hanxu Hou ; Bing Zhu ; Tai Zhou ; Lijia Lu ; Yumeng Zhang</i>	
PARITY DECLUSTERING FOR FAULT-TOLERANT STORAGE SYSTEMS VIA T-DESIGNS	1344
<i>Son Hoang Dau ; Yan Jia ; Chao Jin ; Weiya Xi ; Kheong Sann Chan</i>	
AN EFFICIENT SCHEME TO ENSURE DATA AVAILABILITY FOR A CLOUD SERVICE PROVIDER.....	1352
<i>Seungmin Kang ; Veeravalli, B. ; Khin Mi Mi Aung ; Chao Jin</i>	
A C LIBRARY OF REPAIR-EFFICIENT ERASURE CODES FOR DISTRIBUTED DATA STORAGE SYSTEMS	1358
<i>Chao Tian</i>	
RECT: IMPROVING MAPREDUCE PERFORMANCE UNDER FAILURES WITH RESILIENT CHECKPOINTING TACTICS.....	1364
<i>Hao Wang ; Haopeng Chen ; Fei Hu</i>	
STORE: DATA RECOVERY WITH APPROXIMATE MINIMUM NETWORK BANDWIDTH AND DISK I/O IN DISTRIBUTED STORAGE SYSTEMS.....	1370
<i>Tai Zhou ; Hui Li ; Bing Zhu ; Yumeng Zhang ; Hanxu Hou ; Jun Chen</i>	
PRIVACY-AWARE FILTER-BASED FEATURE SELECTION	1376
<i>Jafer, Y. ; Matwin, S. ; Sokolova, M.</i>	
SECURE DATA STORAGE IN DISTRIBUTED CLOUD ENVIRONMENTS	1381
<i>Jordao, R. ; Aymore Martins, V. ; Buiati, F. ; de Sousa, R.T. ; de Deus, F.E.</i>	

VOLUME 3

LOCATION PREDICTION ATTACKS USING TENSOR FACTORIZATION AND OPTIMAL DEFENSES	1388
<i>Murakami, T. ; Watanabe, H.</i>	

INFORMATION GATEWAY FOR INTEGRATED PHARMACOGENOMICS DATA- IGIPD	1397
<i>Kumar, P. ; Chintalapati, J. ; Neeharika, N. ; Saluja, P. ; Mangala, N. ; Prahlada Rao, B.B.</i>	
PREDICTING A BIOLOGICAL RESPONSE OF MOLECULES FROM THEIR CHEMICAL PROPERTIES USING DIVERSE AND OPTIMIZED ENSEMBLES OF STOCHASTIC GRADIENT BOOSTING MACHINE	1406
<i>Abdunabi, T. ; Basir, O.</i>	
UNDERSTANDING THE EFFECTS OF CONCUSSION USING BIG DATA	1414
<i>Caban, J.J. ; Riedy, G. ; Oakes, T.R. ; Grammer, G. ; DeGraba, T.</i>	
DUPLICATE DRUG DISCOVERY USING HADOOP.....	1420
<i>Shao Hua Cheng ; Yu Shian Chiu ; Shih Yao Dai ; Hui-I Hsiao</i>	
TOWARDS INTEGRATING THE DETECTION OF GENETIC VARIANTS INTO AN IN-MEMORY DATABASE	1423
<i>Fahnrich, C. ; Schapranow, M.P. ; Plattner, H.</i>	
A GENERAL SUPERVISED APPROACH TO SEGMENTATION OF CLINICAL TEXTS.....	1429
<i>Ganesan, K. ; Subotin, M.</i>	
A FAST AND MEMORY-EFFICIENT ALGORITHM FOR LEARNING AND RETRIEVAL OF PHENOTYPIC DYNAMICS IN MULTIVARIATE COHORT TIME SERIES.....	1437
<i>Nemati, S. ; Ghassemi, M.M.</i>	
BIG DATA IN GENOMICS: AN OVERVIEW	1441
<i>Bhardwaj, R. ; Sethi, A. ; Nambiar, R.</i>	
PROTECTIVE EFFECTS OF RHEUMATOID ARTHRITIS IN SEPTIC ICU PATIENTS.....	1446
<i>Sheth, M.B. ; Mark, R. ; Chahin, A. ; Markuzon, N.</i>	
WORKLOAD CHARACTERIZATION FOR MG-RAST METAGENOMIC DATA ANALYTICS SERVICE IN THE CLOUD	1452
<i>Wei Tang ; Bischof, J. ; Desai, N. ; Mahadik, K. ; Gerlach, W. ; Harrison, T. ; Wilke, A. ; Meyer, F.</i>	
TIDE: INTER-CHROMOSOMAL TRANSLOCATION AND INSERTION DETECTION USING EMBEDDINGS.....	1460
<i>Vetro, R. ; Farhoodi, R. ; Kotla, R. ; Haspel, N. ; Weisman, D. ; Rosen, J. ; Simovici, D.</i>	
LOW REDUNDANCY FEATURE SELECTION WITH GROUPED VARIABLES AND ITS APPLICATION TO HEALTHCARE DATA	1467
<i>Hang Wu ; Ji-Jiang Yang ; Jianqiang Li</i>	
PHARMACOLOGICAL CLASS DATA REPRESENTATION IN THE WEB ONTOLOGY LANGUAGE (OWL).....	1473
<i>Qian Zhu ; Cui Tao</i>	
SPATIOTEMPORAL INDEXING TECHNIQUES FOR EFFICIENTLY MINING SPATIOTEMPORAL CO-OCCURRENCE PATTERNS.....	1481
<i>Aydin, B. ; Kempton, D. ; Akkineni, V. ; Gopavaram, S.R. ; Pillai, K.G. ; Angryk, R.</i>	
SCALABLE SOLAR IMAGE RETRIEVAL WITH LUCENE	1491
<i>Banda, J.M. ; Angryk, R.A.</i>	
STREAM MINING FOR SOLAR PHYSICS: APPLICATIONS AND IMPLICATIONS FOR BIG SOLAR DATA	1498
<i>Battams, K.</i>	
A COMPUTER VISION APPROACH TO MINING BIG SOLAR DATA	1507
<i>Felix, S. ; Csillaghy, A.</i>	
ITERATIVE REFINEMENT OF MULTIPLE TARGETS TRACKING OF SOLAR EVENTS	1516
<i>Kempton, D. ; Pillai, K.G. ; Angryk, R.</i>	
IMPROVED DATA EXPLOITATION FOR DKIST HIGH-RESOLUTION SOLAR OBSERVATIONS	1525
<i>Reardon, K.P. ; Berukoff, S.</i>	
MASSIVE LABELED SOLAR IMAGE DATA BENCHMARKS FOR AUTOMATED FEATURE RECOGNITION.....	1533
<i>Schuh, M.A. ; Angryk, R.A.</i>	
SPATIAL DATA ANALYSIS OF COMPLEX URBAN SYSTEMS.....	1541
<i>Peiravian, Farideddin ; Kermanshah, Amirhassan ; Derrible, Sybil</i>	
HUMAN ACTIVITY RECOGNITION IN BIG DATA SMART HOME CONTEXT.....	1547
<i>Azzi, S. ; Dallaire, C. ; Bouzouane, A. ; Bouchard, B. ; Giroux, S.</i>	
A VISUALIZED DATA ANALYSIS FOR BOGUS BUSINESS ENTITY DETECTION	1555
<i>Cheng, H. ; Yi-Chuan Lu ; Chih-Cheng Hsu</i>	
ADVANCED PLANNING AND CONTROL OF MANUFACTURING PROCESSES IN STEEL INDUSTRY THROUGH BIG DATA ANALYTICS: CASE STUDY AND ARCHITECTURE PROPOSAL	1562
<i>Krumeich, J. ; Werth, D. ; Loos, P. ; Schimmelpfennig, J. ; Jacobi, S.</i>	

AN OPEN SCHEMA FOR XML DATA IN HIVE.....	1571
<i>Wuheng Luo ; Bo Liu ; Watfa, A.K.</i>	
PARALLEL AND QUANTITATIVE SEQUENTIAL PATTERN MINING FOR LARGE-SCALE INTERVAL-BASED TEMPORAL DATA.....	1578
<i>Guangchen Ruan ; Hui Zhang ; Plale, B.</i>	
A BIG DATA AGGREGATION, ANALYSIS AND EXPLOITATION INTEGRATED PLATFORM FOR INCREASING SOCIAL MANAGEMENT INTELLIGENCE.....	1586
<i>Sand, G. ; Tsitouras, L. ; Dimitrakopoulos, G. ; Chatzigiannakis, V.</i>	
INVESTIGATING THE ACCURACY OF THE OPENFDA API USING THE FDA ADVERSE EVENT REPORTING SYSTEM (FAERS).....	1594
<i>Shin, J.</i>	
A CCG VIRTUAL SYSTEM FOR BIG DATA APPLICATION COMMUNICATION COSTS ANALYSIS.....	1600
<i>Yongen Yu ; Hongbo Zou ; Wei Tang ; Liwei Liu</i>	
HIGH-FREQUENCY FINANCIAL STATISTICS WITH PARALLEL R AND INTEL XEON PHI COPROCESSOR.....	1607
<i>Jian Zou ; Hui Zhang</i>	
SITUATION AWARE COMPUTING FOR BIG DATA.....	1616
<i>Chan, E.S. ; Gawlick, D. ; Ghoneimy, A. ; Zhen Hua Liu</i>	
TOPIC-SPECIFIC POST IDENTIFICATION IN MICROBLOG STREAMS.....	1622
<i>Karunasekera, S. ; Harwood, A. ; Samarawickrama, S. ; Ramamohanarao, K. ; Robins, G.</i>	
HANDLING SMART ENVIRONMENT DEVICES, DATA AND SERVICES AT THE SEMANTIC LEVEL WITH THE FI-WARE CORE PLATFORM	1629
<i>Ramparany, F. ; Galan Marquez, F. ; Soriano, J. ; Elsaleh, T.</i>	
AN IOT/IOE ENABLED ARCHITECTURE FRAMEWORK FOR PRECISION ON SHELF AVAILABILITY: ENHANCING PROACTIVE SHOPPER EXPERIENCE	1636
<i>Vargheese, R. ; Dahir, H.</i>	
LEARNING MACHINES FOR COMPUTATIONAL EPIDEMIOLOGY.....	1642
<i>Boman, M. ; Gillblad, D.</i>	
EPIDEMIOLOGICAL MODELING OF BOVINE BRUCELLOSIS IN INDIA	1647
<i>Kang, G.J. ; Gunaseelan, L. ; Abbas, K.M.</i>	
BIG DATA PROBLEMS ON DISCOVERING AND ANALYZING CAUSAL RELATIONSHIPS IN EPIDEMIOLOGICAL DATA.....	1652
<i>Yiheng Liang ; Mikler, A.R.</i>	
SPATIAL BIG DATA ANALYTICS OF INFLUENZA EPIDEMIC IN VELLORE, INDIA.....	1660
<i>Lopez, D. ; Gunasekaran, M. ; Murugan, B.S. ; Kaur, H. ; Abbas, K.M.</i>	
MULTIWAY ANALYSIS OF BRIDGE STRUCTURAL TYPES IN THE NATIONAL BRIDGE INVENTORY (NBI): A TENSOR DECOMPOSITION APPROACH.....	1666
<i>Adarkwa, O. ; Schumacher, T. ; Attoh-Okine, N.</i>	
BIG DATA CHALLENGES IN RAILWAY ENGINEERING.....	1672
<i>Attoh-Okine, N.</i>	
EFFICIENT TRAFFIC SPEED FORECASTING BASED ON MASSIVE HETEROGENOUS HISTORICAL DATA	1675
<i>Xing-Yu Chen ; Hsing-Kuo Pao ; Yuh-Jye Lee</i>	
TOPOLOGICAL MODELS OF DOCUMENT-QUERY SETS IN RETRIEVAL FOR ENTERPRISE INFORMATION MANAGEMENT.....	1683
<i>Deolalikar, V.</i>	
A DYNAMIC PROGRAMMING APPROACH FOR 4D FLIGHT ROUTE OPTIMIZATION	1689
<i>Kiss-Toth, C. ; Takacs, G.</i>	
IMPACT ANALYSIS OF EXTREME EVENTS ON FLOWS IN SPATIAL NETWORKS.....	1694
<i>Kermanshah, A. ; Karduni, A. ; Peiravian, F. ; Derrible, S.</i>	
APPLICATIONS OF LINKED DATA IN THE RAIL DOMAIN	1700
<i>Morris, C. ; Easton, J. ; Roberts, C.</i>	
METAHEURISTICS IN BIG DATA: AN APPROACH TO RAILWAY ENGINEERING.....	1707
<i>Nunez, S.G. ; Attoh-Okine, N.</i>	
FACILITATING MAINTENANCE DECISIONS ON THE DUTCH RAILWAYS USING BIG DATA: THE ABA CASE STUDY	1713
<i>Nunez, A. ; Hendriks, J. ; Zili Li ; De Schutter, B. ; Dollevoet, R.</i>	
SPATIAL DATA ANALYSIS OF COMPLEX URBAN SYSTEMS.....	1719
<i>Peiravian, F. ; Kermanshah, A. ; Derrible, S.</i>	

EVALUATING STRUCTURAL ENGINEERING FINITE ELEMENT ANALYSIS DATA USING MULTIWAY ANALYSIS	1725
<i>Radovic, M. ; McConnell, J.</i>	
MULTI-OBJECTIVE OPTIMIZATION FOR RESILIENT AIRLINE NETWORKS USING SOCIOECONOMIC-ENVIRONMENTAL DATA	1733
<i>Sawai, H. ; Sato, A.-H.</i>	
PREDICTING FLIGHT ARRIVAL TIMES WITH A MULTISTAGE MODEL.....	1743
<i>Takacs, G.</i>	
ONTOLOGY-DRIVEN DATA INTEGRATION FOR RAILWAY ASSET MONITORING APPLICATIONS	1750
<i>Tutcher, J.</i>	
SOME EXAMPLES OF BIG DATA IN RAILROAD ENGINEERING	1761
<i>Zarembski, A.M.</i>	
TAKING AN ELECTRONIC TICKETING SYSTEM TO THE CLOUD: DESIGN AND DISCUSSION.....	1768
<i>Araujo, F. ; Curado, M. ; Furtado, P. ; Barbosa, R.</i>	
A CONTENTION AWARE HYBRID EVALUATOR FOR SCHEDULERS OF BIG DATA APPLICATIONS IN COMPUTER CLUSTERS.....	1778
<i>Bardhan, S. ; Menasce, D.A.</i>	
RULEMR: CLASSIFICATION RULE DISCOVERY WITH MAPREDUCE.....	1787
<i>Kolias, V. ; Kolias, C. ; Anagnostopoulos, I. ; Kayafas, E.</i>	
A RELATIONAL DATABASE SCHEMA ON THE TRANSACTIONAL KEY-VALUE STORE SCALARIS	1796
<i>Kruber, N. ; Schintke, F. ; Berlin, M.</i>	
COMMUNITY STRUCTURE ANALYSIS IN BIG CLIMATE DATA.....	1805
<i>McGuire, M.P. ; Nguyen, N.P.</i>	
THE BEST OF TWO WORLDS: INTEGRATING IBM INFOSPHERE STREAMS WITH APACHE YARN	1814
<i>Nabi, Z. ; Wagle, R. ; Bouillet, E.</i>	
TEMPORAL BIPARTITE PROJECTION AND LINK PREDICTION FOR ONLINE SOCIAL NETWORKS.....	1819
<i>Tsunghan Wu ; Sheau-Harn Yu ; Wanjiun Liao ; Cheng-Shang Chang</i>	
THE DEEP FILM ACCESS PROJECT: ONTOLOGY AND METADATA DESIGN FOR DIGITAL FILM PRODUCTION ASSETS.....	1827
<i>Atkinson, S. ; Lehmann, J. ; Evans, R.</i>	
MINING MICRODATA: ECONOMIC OPPORTUNITY AND SPATIAL MOBILITY IN BRITAIN AND THE UNITED STATES, 1850–1881.....	1831
<i>Baskerville, P. ; Dillon, L. ; Inwood, K. ; Roberts, E. ; Ruggles, S. ; Schurer, K. ; Warren, J.R.</i>	
MINING MOBILE YOUTH CULTURES.....	1840
<i>Blanke, T. ; Greenway, G. ; Pybus, J. ; Cote, M.</i>	
SCIENTIFIC FINDINGS AS BIG DATA FOR RESEARCH SYNTHESIS: THE METABUS PROJECT	1844
<i>Bosco, F. ; Uggerslev, K. ; Steel, P.</i>	
SCALING HISTORICAL TEXT RE-USE	1849
<i>Buchler, M. ; Franzini, G. ; Franzini, E. ; Moritz, M.</i>	
REVOLUTIONARY ENTITIES: TURNING DATA INTO KNOWLEDGE TO DRIVE PERSONALIZED EXPLORATION OF THE IRISH RISING OF 1916.....	1858
<i>Conlan, O. ; O'Connor, A. ; Loinsigh, O.N. ; Munnelly, G. ; Lawless, S. ; Murphy, R.</i>	
UNDERSTANDING THE ROLE OF MEDICAL EXPERTS DURING A PUBLIC HEALTH CRISIS DIGITAL TOOLS AND LIBRARY RESOURCES FOR RESEARCH ON THE 1918 SPANISH INFLUENZA.....	1865
<i>Ewing, E.T. ; Gad, S. ; Ramakrishnan, N. ; Reznick, J.S.</i>	
A METADATA INFRASTRUCTURE FOR THE ANALYSIS OF PARLIAMENTARY PROCEEDINGS.....	1873
<i>Gartner, R.</i>	
SCALED ENTITY SEARCH: A METHOD FOR MEDIA HISTORIOGRAPHY AND RESPONSE TO CRITIQUES OF BIG HUMANITIES DATA RESEARCH	1877
<i>Hoyt, E. ; Hughes, K. ; Long, D. ; Tran, A. ; Ponto, K.</i>	
ON THE COVERAGE OF SCIENCE IN THE MEDIA: A BIG DATA STUDY ON THE IMPACT OF THE FUKUSHIMA DISASTER.....	1886
<i>Lansdall-Welfare, T. ; Sudhahar, S. ; Veltri, G.A. ; Cristianini, N.</i>	

INTEGRATING DATA MINING AND DATA MANAGEMENT TECHNOLOGIES FOR SCHOLARLY INQUIRY	1893
<i>Larson, R.R. ; Marciano, R. ; Chien-Yi Hou ; Shreyas ; Watry, P. ; Harrison, J. ; Aguilar, L. ; Fuselier, J.</i>	
THE EXCEPTIONAL AND THE EVERYDAY: 144 HOURS IN KIEV	1898
<i>Manovich, L. ; Tifentale, A. ; Yazdani, M. ; Chow, J.</i>	
DEALING WITH HETEROGENEOUS BIG DATA WHEN GEOPARSING HISTORICAL CORPORA	1906
<i>Rupp, C.J. ; Rayson, P. ; Gregory, I. ; Hardie, A. ; Joulain, A. ; Hartmann, D.</i>	
BIGEXCEL: A WEB-BASED FRAMEWORK FOR EXPLORING BIG DATA IN SOCIAL SCIENCES	1910
<i>Saleem, M.A. ; Varghese, B. ; Barker, A.</i>	
PROBABILISTIC ESTIMATES OF ATTRIBUTE STATISTICS AND MATCH LIKELIHOOD FOR PEOPLE ENTITY RESOLUTION	1918
<i>Xin Wang ; Ang Sun ; Kardes, H. ; Agrawal, S. ; Lin Chen ; Borthwick, A.</i>	
A COMPUTATIONAL PIPELINE FOR CROWDSOURCED TRANSCRIPTIONS OF ANCIENT GREEK PAPYRUS FRAGMENTS	1926
<i>Williams, A.C. ; Wallin, J.F. ; Haoyu Yu ; Perale, M. ; Carroll, H.D. ; Lamblin, A.-F. ; Fortson, L. ; Obbink, D. ; Lintott, C.J. ; Brusuelas, J.H.</i>	
“SUSTAINABLE ASSEMBLAGE FOR ENERGY (SAE)” INSIDE INTELLIGENT URBAN AREAS: HOW MASSIVE HETEROGENEOUS DATA COULD HELP TO REDUCE ENERGY FOOTPRINTS AND PROMOTE SUSTAINABLE PRACTICES AND AN ECOLOGICAL TRANSITION	1932
<i>Calvez, P. ; Soulier, E.</i>	
A MODEL ARCHITECTURE FOR BIG DATA APPLICATIONS USING RELATIONAL DATABASES	1940
<i>Durham, E.-E.A. ; Rosen, A. ; Harrison, R.W.</i>	
OPTIMIZING GRAPH QUERIES WITH GRAPH JOINS AND SPRINKLE SPARQL	1948
<i>Goodman, E.L. ; Jimenez, E. ; al-Saffar, S. ; Joslyn, C. ; Haglin, D. ; Grunwald, D.</i>	
VESSEL ROUTE ANOMALY DETECTION WITH HADOOP MAPREDUCE	1956
<i>Xiaoguang Wang ; Xuan Liu ; Bo Liu ; de Souza, E.N. ; Matwin, S.</i>	
HBGSIM: A STRUCTURAL SIMILARITY MEASUREMENT OVER HETEROGENEOUS BIG GRAPHS	1962
<i>Jiazhen Nian ; Shan Jiang ; Yan Zhang</i>	
KNOWLEDGE BASED DIMENSIONALITY REDUCTION FOR TECHNICAL TEXT MINING	1970
<i>Shalaby, W. ; Zadrozny, W. ; Gallagher, S.</i>	
A DISTRIBUTED INSTANCE-WEIGHTED SVM ALGORITHM ON LARGE-SCALE IMBALANCED DATASETS	1976
<i>Xiaoguang Wang ; Xuan Liu ; Matwin, S.</i>	
A LAYER BASED ARCHITECTURE FOR PROVENANCE IN BIG DATA	1983
<i>Agrawal, R. ; Imran, A. ; Seay, C. ; Walker, J.</i>	
SHARING BEST PRACTICES FOR THE IMPLEMENTATION OF BIG DATA APPLICATIONS IN GOVERNMENT AND SCIENCE COMMUNITIES	1990
<i>Aron, J.L. ; Niemann, B.</i>	
BIG DATA: CHALLENGES, PRACTICES AND TECHNOLOGIES: NIST BIG DATA PUBLIC WORKING GROUP WORKSHOP AT IEEE BIG DATA 2014	1993
<i>Grady, N.W. ; Underwood, M. ; Roy, A. ; Chang, W.L.</i>	
BIG DATA MACHINE LEARNING AND GRAPH ANALYTICS: CURRENT STATE AND FUTURE CHALLENGES	1998
<i>Huang, H.H. ; Hang Liu</i>	
A STANDARD FOR BENCHMARKING BIG DATA SYSTEMS	2000
<i>Nambiar, R.</i>	
ADDRESSING DATA VERACITY IN BIG DATA APPLICATIONS	2003
<i>Aman, S. ; Chelms, C. ; Prasanna, V.</i>	
MACHINE LEARNING AND INTERACTIVE VISUALIZATION APPLIED TO TB-SIZED IMAGES OF STEM CELLS	2006
<i>Amelot, J. ; Bajcsy, P. ; Plant, A. ; Brady, M.</i>	
BIG DATA: A NEW CHALLENGE FOR TOURISM	2007
<i>Chareyron, G. ; Da-Rugna, J. ; Raimbault, T.</i>	
OME: TOOL FOR GENERATING AND MANAGING METADATA TO HANDLE BIGDATA	2010
<i>Devarakonda, R. ; Shrestha, B. ; Palanisamy, G. ; Hook, L. ; Killeffer, T. ; Krassovski, M. ; Boden, T. ; Cook, R. ; Zolly, L. ; Hutchison, V. ; Frame, M. ; Cialella, A. ; Lazer, K.</i>	

THE EMBERS ARCHITECTURE FOR STREAMING PREDICTIVE ANALYTICS	2013
<i>Doyle, A. ; Katz, G. ; Summers, K. ; Ackermann, C. ; Zavorin, I. ; Zunsik Lim ; Muthiah, S. ; Liang Zhao ; Chang-Tien Lu ; Butler, P. ; Khandpur, R.P. ; Fayed, Y. ; Ramakrishnan, N.</i>	
A NOVEL APPROACH TO DETERMINE DOCKING LOCATIONS USING FUZZY LOGIC AND SHAPE DETERMINATION	2016
<i>Umoja, C. ; Torrance, J.T. ; Durham, E.-E.A. ; Rosen, A. ; Harrison, R.W.</i>	
LINKED OPEN DATA MINING FOR DEMOCRATIZATION OF BIG DATA.....	2019
<i>Espinosa, R. ; Garriga, L. ; Zubcoff, J.J. ; Mazon, J.-N.</i>	
BUILDING WRANGLER: A TRANSFORMATIONAL DATA INTENSIVE RESOURCE FOR THE OPEN SCIENCE COMMUNITY	2022
<i>Gaffney, N. ; Jordan, C. ; Minyard, T. ; Stanzione, D.</i>	
CELAR: AUTOMATED APPLICATION ELASTICITY PLATFORM	2025
<i>Giannakopoulos, I. ; Papailiou, N. ; Mantas, C. ; Konstantinou, I. ; Tsoumakos, D. ; Koziris, N.</i>	
SEMANTIC HMC FOR BIG DATA ANALYSIS	2028
<i>Hassan, T. ; Peixoto, R. ; Cruz, C. ; Bertaux, A. ; Silva, N.</i>	
A LAYER BASED ARCHITECTURE FOR PROVENANCE IN BIG DATA.....	2031
<i>Imran, Ashiq ; Agrawal, Rajeev ; Walker, Jessie ; Gomes, Anthony</i>	
B-DIDS: MINING ANOMALIES IN A BIG-DISTRIBUTED INTRUSION DETECTION SYSTEM	2034
<i>Janeja, V.P. ; Azari, A. ; Namayanja, J.M. ; Heilig, B.</i>	
ENABLING GENOMIC ANALYSIS ON FEDERATED CLOUDS	2037
<i>Fan Jiang ; Shoffner, M. ; Castillo, C. ; Schmitt, C.</i>	
CHALLENGES OF DATA INTEGRATION AND INTEROPERABILITY IN BIG DATA	2040
<i>Kadadi, A. ; Agrawal, R. ; Nyamful, C. ; Atiq, R.</i>	
LARGE SCALE AUTHOR NAME DISAMBIGUATION IN DIGITAL LIBRARIES	2043
<i>Khabsa, M. ; Treeratpituk, P. ; Giles, C.L.</i>	
REAL-TIME TRAFFIC INCIDENT DETECTION USING PROBE-CAR DATA ON THE TOKYO METROPOLITAN EXPRESSWAY.....	2045
<i>Kinoshita, A. ; Takasu, A. ; Adachi, J.</i>	
DIFFERENTIALLY PRIVATE MODELS OF TOLLGATE USAGE: THE MILAN TOLLGATE DATA SET	2048
<i>Manfredi, N. ; Mir, D.J. ; Lu, S. ; Sanchez, D.</i>	
MODISSENSE: A DISTRIBUTED PLATFORM FOR SOCIAL NETWORKING SERVICES OVER MOBILE DEVICES	2051
<i>Mytilinis, I. ; Giannakopoulos, I. ; Konstantinou, I. ; Doka, K. ; Koziris, N.</i>	
A CHALLENGE OF AUTHORSHIP IDENTIFICATION FOR TEN-THOUSAND-SCALE MICROBLOG USERS.....	2054
<i>Okuno, S. ; Asai, H. ; Yamana, H.</i>	
COGNITIVE MAP OF TOURIST BEHAVIOR BASED ON TRIPADVISOR	2057
<i>Raimbault, T. ; Chareyron, G. ; Krzyzanowski-Guillot, C.</i>	
BICLUSTERING USING SPARK-MAPREDUCE	2060
<i>Sarazin, T. ; Lebbah, M. ; Azzag, H.</i>	
A SUMMARIZATION PARADIGM FOR BIG DATA.....	2063
<i>Shah, Z. ; Mahmood, A.N.</i>	
AN OPEN SOURCE FRAMEWORK TO ADD SPATIAL EXTENT AND GEOSPATIAL VISIBILITY TO BIG DATA.....	2066
<i>Shrestha, B. ; Devarakonda, R. ; Palanisamy, G.</i>	
DEVELOPING A CLOUD COMPUTING PLATFORM FOR BIG DATA: THE OPENSTACK NOVA CASE.....	2069
<i>Teixeira, J.</i>	
LANGUAGE, CULTURAL INFLUENCES AND INTELLIGENCE IN HISTORICAL GAZETTEERS OF THE GREAT WAR.....	2072
<i>Warren, R. ; Bo Liu</i>	
THE BOT WILL SERVE YOU NOW: AUTOMATING ACCESS TO ARCHIVAL MATERIALS.....	2075
<i>Westgard, J.A.</i>	
INCREMENTAL AND PARALLEL SPATIAL ASSOCIATION MINING	2077
<i>Jin Soung Yoo ; Boulware, D.</i>	
SHARDING FOR LITERATURE SEARCH VIA CUTTING CITATION GRAPHS.....	2079
<i>Haozhen Zhao</i>	
REPAIR EFFICIENT STORAGE CODES VIA COMBINATORIAL CONFIGURATIONS	2082
<i>Bing Zhu ; Hui Li ; Shum, K.W.</i>	
Author Index	