2017 IEEE International Congress on Big Data
(BigData Congress 2017)

Honolulu, Hawaii, USA
25-30 June 2017
2017 IEEE 6th International Congress on Big Data

BigDataCongress 2017

Table of Contents

Message from the General Chairs .................................................................xiii
Message from the Program Chairs ................................................................xv
Organizing Committee ..................................................................................xvi
Technical Program Committee ......................................................................xviii
IEEE Computer Society Technical Committee on Services Computing Services Society .................................................................xxii

BigData Congress Research Track

Research Session 1: Machine Learning for Big Data

Comprehensive Predictions of Tourists’ Next Visit Location Based on Call Detail Records Using Machine Learning and Deep Learning Methods ........................................1
Nai Chun Chen, Wanqin Xie, Roy E. Welsch, Kent Larson, and Jenny Xie

Bayesian Networks with Structural Restrictions: Parallelization, Performance, and Efficient Cross-Validation .......................................................................................7
Hao Peng, Zhe Jin, and John A. Miller

Building an SVM Classifier for Automated Selection of Big Data ..................15
Junhua Ding, Jiabin Wang, Xiaojun Kang, and Xin-Hua Hu

Research Session 2: Big Data Analytics Framework

Adaptive Concurrent Query Execution Framework for an Analytical In-Memory Database System ........................................................................................................23
Harshad Deshmukh, Hakan Memisoglu, and Jignesh M. Patel

Toward Scalable Indexing and Search on Distributed and Unstructured Data ..........31
Alexandru Iulian Orhean, Itua Ijagbone, Iloa Raicu, Kyle Chard, and Dongfang Zhao

Scalable, Efficient Anonymization with INCOGNITO - Framework & Algorithm ......................................................39
Antorweep Chakravorty, Chunming Rong, K.R. Jayaram, and Shu Tao
Research Session 3: Big Data Privacy and Protection

Differentially Private Trajectory Analysis for Points-of-Interest .................................................................49
   Chao Li, Balaji Palanisamy, and James Joshi
Privacy-Preserving Inductive Learning with Decision Trees ...........................................................................57
   Stacey Truex, Ling Liu, Mehmet Emre Gursoy, and Lei Yu
SafeNVM: A Non-Volatile Memory Store with Thread-Level Page Protection ........................................65
   Pradeep Kumar and H. Howie Huang

Research Session 4: Social Big Data

Maximizing the Effectiveness of Advertising Campaigns on Twitter ...............................................................73
   Yan Mei, Weiliang Zhao, and Jian Yang
Big Social Data Analytics in Football: Predicting Spectators and TV Ratings from Facebook Data ..................81
   Nicolai H. Egebjerg, Niklas Hedegaard, Gerda Kuum,
   Raghava Rao Mukkamala, and Ravi Vatrapu
Big Social Data Analytics for Public Health: Predicting Facebook Post Performance Using Artificial Neural Networks and Deep Learning ..........................................................89
   Nadiya Straton, Raghava Rao Mukkamala, and Ravi Vatrapu

Research Session 5: Big Data Systems

An Integrated Software System for Supporting Real-Time Near-Infrared Spectral Big Data Analysis and Management ..............................................................................................................97
   Liping Zhao, Shupeng Hu, Xiaojun Zeng, Yuejin Wu, Yanqing Lin, Jing Liu,
   and Shuang Fan
A Model-Driven Methodology for Big Data Analytics-as-a-Service ...............................................................105
   Claudio A. Ardagna, Valerio Bellandi, Paolo Ceravolo, Ernesto Damiani,
   Michele Bezzi, and Cedric Hebert
Bleach: A Distributed Stream Data Cleaning System ...................................................................................113
   Yongchao Tian, Pietro Michiardi, and Marko Vukolić

Research Session 6: Big Data Performance

Zero-Day Attack Identification in Streaming Data Using Semantics and Spark ............................................121
   Sai C. Pallaprolu, Rishi Sankineni, Muthukumar Thevar, George Karabatis,
   and Jianwu Wang
Big Data Processing: Scalability with Extreme Single-Node Performance ........................................... 129

Venkatraman Govindaraju, Sam Idicula, Sandeep Agrawal,
Venkatanathan Vardarajan, Arun Raghavan, Jarod Wen, Cagri Balkesen,
Georgios Giannikis, Nipun Agarwal, and Eric Sedlar

Parallel Compression and Indexing of Large-Scale Geospatial Raster Data
with GPGPUs ............................................................................................................................................ 137

Nathalie Kaligirwa, Eleazar Leal, Le Gruenwald, Jianting Zhang, and Simin You

Research Session 7: Resource Management in Big Data

Scheduling Recurring Distributed Dataflow Jobs Based on Resource
Utilization and Interference ...................................................................................................................... 145

Lauritz Thamsen, Benjamin Rabier, Florian Schmidt, Thomas Renner,
and Odej Kao

Resource Allocation in Cloud-Based Distributed Cameras ................................................................. 153

Bikash Agrawal, Jayachander Surbiryala, and Chunming Rong

Resource Management for Parallel Processing Frameworks with Load
Awareness at Worker Side .................................................................................................................. 161

Son-Hai Ha, Patrick Brown, and Pietro Michiardi

Research Session 8: Big Data Processing

CEPaaS: Complex Event Processing as a Service ................................................................................ 169

Wilson A. Higashino, Miriam A.M. Capretz, and Luiz F. Bittencourt

Towards Elucidating the Structural Principles of Host-Pathogen
Protein-Protein Interaction Networks: A Bioinformatics Survey .................................................. 177

Huaming Chen, Jiangning Song, Geng Sun, Jun Shen, and Lei Wang

OB-Tree: Accelerating Data Cleaning in Out-of-Core Column-Store Databases ........................ 185

Feng Yu, Brandon J. Latronics, Tyler Matacic, and Eric S. Jones

Research Session 9: Big Data Learning

A Deep-Learning Based Precipitation Forecasting Approach Using Multiple
Environmental Factors ......................................................................................................................... 193

Pengcheng Zhang, Lei Zhang, Hareton Leung, and Jimin Wang

What Are My Users Looking for When Preparing a Big Data Campaign ...................................... 201

Claudio Ardagna, Paolo Ceravolo, Guido Lena Cota,
Muhammad Muneeb Kiani, and Ernesto Damiani

Content Embedding Regularized Matrix Factorization for Recommender
Systems ............................................................................................................................................... 209

Hao Wu, Zhengxin Zhang, Kun Yue, Binbin Zhang, and Ruichao Zhu
Research Session 10: Big Data Case Study

A Spatio-Temporal Mining Approach for Enhancing Satellite Data Availability:
A Case Study on Blue Green Algae ..............................................................216
  Vinay Boddula, Lakshmish Ramaswamy, and Deepak Mishra

Real-Time Anomaly Detection of Short-Time-Scale GWAC Survey Light Curves ..............................................................224
  Tianzhi Feng, Zhihui Du, Yankui Sun, Jianyan Wei, Jing Bi, and Jason Liu

BigData Congress Visionary Track

Parallel Processing over Spatial-Temporal Datasets from Geo, Bio, Climate and Social Science Communities: A Research Roadmap .......................................................232
  Sushil K. Prasad, Danial Aghajarian, Michael McDermott, Dhara Shah, Mohamed Mokbel, Satish Puri, Sergio J. Rey, Shashi Shekhar, Yiqun Xe, Ranga Raju Vatsavai, Fusheng Wang, Yanhui Liang, Hoang Vo, and Shaowen Wang

Adding Support for Theory in Open Science Big Data ..............................................................251
  John A. Miller, Hao Peng, and Michael E. Cotterell

Towards a Framework for Developing Cyber Privacy Metrics: A Vision Paper ..............................................................256
  Bhavani Thuraisingham, Murat Kantarcioglu, Elisa Bertino, and Chris Clifton

BigData Congress Application Track

Session 1: Big Data QoS

De-Centralized Reputation-Based Trust Model to Discriminate between Cloud Providers Capable of Processing Big Data ..............................................................266
  Hadeel T. El Kassabi and Mohamed Adel Serhani

A Multivariate Clustering Approach for Infrastructure Failure Predictions ..............................................................274
  Simon Luo, Victor W. Chu, Jianlong Zhou, Fang Chen, Raymond K. Wong, and Weidong Huang

Microblog Sentiment Classification Using Parallel SVM in Apache Spark ..............................................................282
  Bo Yan, Zijiang Yang, Yitian Ren, Xing Tan, and Eric Liu

Session 2: Machine Learning in Big Data

Drowsy Driving Warning System Based on GS1 Standards with Machine Learning ..............................................................289
  Seongjoo Moon, Moran Min, Jangwoo Nam, Jaeyou Park, Doheon Lee, and Daeyoung Kim
An Effective Distributed GHSOM Algorithm for Unsupervised Clustering on Big Data .................................................................297
  Chui-Hui Chiu, Jin-Jie Chen, and Fang Yu

Integrating Extractive and Abstractive Models for Long Text Summarization .................................................................305
  Shuai Wang, Xiang Zhao, Bo Li, Bin Ge, and Daquan Tang

Session 3: Big Data Processing

SCMAT: A Mechanism Presuming SCMs to Efficiently Enable Both OLAP and OLTP .................................................................313
  Takamitsu Shioi, Kenji Hatano, and Haruo Yokota

A Real-Time Scheduling Strategy Based on Processing Framework of Hadoop .................................................................321
  Fangbing Chen, Ji Liu, and Yuesheng Zhu

Developing a Real-Time Data Analytics Framework for Twitter Streaming Data .................................................................329
  Babak Yadranjiaghdam, Seyedfaraz Yasrobi, and Nasseh Tabrizi

Session 4: Big Data Mining

Locating Sources in Online Social Networks via Random Walk .................................................................337
  Zhijian Zhang, Kun Yue, Zhengbao Sun, Ling Liu, and Weiyi Liu

Big Data and Analysis of Data Transfers for International Research Networks Using NetSage .................................................................344
  Alberto Gonzalez, Jason Leigh, Sean Peisert, Brian Tierney, Edward Balas, Predrag Radulovic, and Jennifer M. Schopf

PAS4SCF - A Publication Analytics System for Services Conference Federation (SCF 2004-2016) .................................................................352
  Liang-Jie Zhang and Huan Chen

Session 5: Big Data Algorithms

I-Sampling: A New Block-Based Sampling Method for Large-Scale Dataset .................................................................360
  Yulin He, Joshua Zhexue Huang, Hao Long, Qiang Wang, and Chenghao Wei

Leveraging Stacked Denoising Autoencoder in Prediction of Pathogen-Host Protein-Protein Interactions .................................................................368
  Huaming Chen, Jun Shen, Lei Wang, and Jiangning Song

Experiences with the Twitter Health Surveillance (THS) System .................................................................376
  Manuel Rodriguez-Martinez
Session 6: Big Data Case Studies

Data Governance Framework for Big Data Implementation with a Case of Korea .......................................................................................................................................................... 384
  Hee Young Kim and June-Suh Cho

Educational and Scientific Applications of Climate Model Diagnostic Analyzer ................................................................. 392
  Seungwon Lee, Chengxing Zhai, Terence Kubar, Benyang Tang, Lei Pan, Julin Li, Jia Zhang, and Qihao Bao

Implementation of Disaster Prevention-Oriented Information Service Platform of XRAIN on DIAS ................................................................................................................................................................................. 398
  Hitomi Sano, Eiji Ikoma, Masaru Kitsuregawa, and Masato Oguchi

Session 7: Big Data Models

New E-Commerce User Interest Patterns .......................................................................................................................................................... 406
  Matthias Volk, Abed Elrahman Shareef, Naoum Jamous, and Klaus Turowski

Big Data Analysis and Visualization for the Smart Grid ............................................................................................................................................. 414
  Alejandro Sanchez and Wilson Rivera

Session 8: Big Data Analysis

Compliance Analysis of Configurable Business Process Model Based on Extend CTL .................................................................................................................................................................................. 419
  Huang Yiwang, Feng Zaiwen, He Keqing, and Xu Songjin

Any-Time Methods for Time-Series Prediction with Missing Observations ........................................................................................................... 427
  Yanjie Wei, Hongzhi Wang, Shengfei Shi, Hong Gao, and Jianzhong Li

BigData Congress Theme Track

Session 1: Big Data Management

TeenRead: An Adolescents Reading Recommendation System Towards Online Bibliotherapy ............................................................................................................................................................................. 431
  Yunxing Xin, Yongqiang Chen, Li Jin, Yici Cai, and Ling Feng

A Resilient Index Graph for Querying Large Biological Scientific Data .................................................................................................................. 435
  Liang Li, Zhihong Shen, Jianhui Li, Dongjiang Liu, Huajin Wang, Lipeng Wang, and Qinglan Sun

SPARQL Query Parallel Processing: A Survey .......................................................................................................................................................... 444
  Jiaying Feng, Chenhong Meng, Jiaming Song, Xiaowang Zhang, Zhiyong Feng, and Lei Zou
Session 2: Big Data Management

Burst-Based Event Classification on Weakly Labeled Time Series Data of Sensors .................................................................452
  Hanbo Zhang, Yawen Wang, Peng Wang, and Wei Wang

Bootstrapping a Blockchain Based Ecosystem for Big Data Exchange .................................................................460
  Jinchuan Chen and Yunzhi Xue

The Performance of SQL-on-Hadoop Systems - An Experimental Study ...............................................................464
  Xiongpai Qin, Yueguo Chen, Jun Chen, Shuai Li, Jiesi Liu, and Huijie Zhang

Session 3: Real-Time Big Data Case Analytics

Real Time Social Attitude Expression Prediction .................................................................472
  Jiawei Zhang, Philip S. Yu, Charu C. Aggarwal, and Limeng Cui

A Hybrid Learning Framework for Imbalanced Stream Classification .................................................................480
  Wenbin Zhang and Jianwu Wang

Personalized Recommendation Algorithm Based on LFM with QoS Constraint .................................................................488
  Ying Jin and Yi-wen Zhang

BigData Congress Short Paper Track

Session 1: Big Data Processing

Clustering Data Streams with Adaptive Forgetting .................................................................494
  Gopi Chand Nutakki and Olfa Nasraoui

Big Data Pre-Processing: Closing the Data Quality Enforcement Loop .................................................................498
  Ikbal Taleb and Mohamed Adel Serhani

Cloud-Based and Big Data-Enabled Brokerage System for Smart Grids .................................................................502
  Howard Martinez Meza and Manuel Rodriguez-Martinez

Graph Computation and Its Applications in Smart Grid .................................................................507
  Guangyi Liu, Kewen Liu, Di Shi, Wendong Zhu, Zhiwei Wang, and Xi Chen

Session 2: QoS of Big Data

Open Medical Big Data and Open Consent and Their Impact on Privacy .................................................................511
  Jingquan Li

Enhancing Energy, Health and Security Policy by Extracting, Enriching and Interfacing Next Generation Data in the Transport Domain (A Study on the Use of Big Data in Cross-Sectoral Policy Development) .................................................................515
  Susan Grant-Muller, Frances Hodgson, Nick Malleson, and Robert Snowball

Managing Emerging Risks and Liabilities in Data-Enabled Solutions .................................................................519
  Olaitan Olaleye, Talmai Oliveira, and Alexandru Darie
Ecosystem Design of Big Data through Previous Study Analysis in the World: Policy Design for Big Data as Public Goods ..................................................................................................................525
   HyeJung Moon, KwanWoo Kim, SungKyung Kim, and JangHo Park

Challenges of Software Testing for Astronomical Big Data .................................................................................................................................529
   Lixiao Zhou and Maohai Huang

**Session 3: Big Data Applications**

A Distributed System for Finding High Profit Areas over Big Taxi Trip Data with MogoDB and Spark ..................................................................................................................533
   Fadhilah Kurnia Putri and Joonho Kwon

Comparison of Gambling Industry through Study Analysis in the World .................................................................................................................................537
   HyeJung Moon, Thomas Yang, SungKyung Kim, and JangHo Park

Big Data Analysis on Gambling Related News in South Korea .................................................................................................................................541
   HyeJung Moon, SungKyung Kim, Thomas Yang, and KwanWoo Kim

Semantic Analysis of Transportation Related Researches .................................................................................................................................545
   SiYeun Lee and HyeJung Moon

**Session 4: Satellite Session**

Business Graphing for Internet-Enabled Enterprises .................................................................................................................................549
   Bo Hu, Jian Wang, Yutao Ma, and Liang-Jie Zhang

An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends .................................................................................................................................557
   Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang

A New Sleep Mechanism Based on SDN Framework .................................................................................................................................565
   Yu Peizhao, Xu Lei, Gu Jinguang, and He Heng

CloudSec: A Novel Approach to Verifying Security Conformance at the Bottom of the Cloud .................................................................................................................................569
   Xiaochen Liu, Chunhe Xia, Tianbo Wang, and Li Zhong

Cutting Parameters Optimization Based on ITLBO Algorithm with Big Data Driven .................................................................................................................................577
   Chen Wang, Yi Wang, KeSheng Wang, Yao Dong, Xueliang Zhou, and Shenghuai Wang

**Author Index** .................................................................................................................................582