2016 IEEE International Congress on Big Data (BigData Congress 2016)

San Francisco, California, USA
27 June – 2 July 2016
2016 IEEE International Congress on Big Data

BigDataCongress 2016

Table of Contents

Message from the General Chairs ........................................................................................................xii
Message from the Program Committee Chairs ..........................................................................................xiv
Organizing Committee ..................................................................................................................................xv
Program Committee ..................................................................................................................................xviii
IEEE Computer Society Technical Committee on Services Computing ..................................................xxv

BigData Congress Research Track

Session 1: Best Paper Session 1

Boosting Vertex-Cut Partitioning for Streaming Graphs .............................................................................1
  Hooman Peiro Sajjad, Amir H. Payberah, Fatemeh Rahimian, Vladimir Vlassov, and Seif Haridi

Optimizing Hadoop Framework for Solid State Drives .............................................................................9
  Jaeki Hong, Liang Li, Chihye Han, Bingxu Jin, Qichao Yang, and Zilong Yang

Continual and Cost-Effective Partitioning of Dynamic Graphs for Optimizing Big Graph Processing Systems ..................................................................................................................18
  Amirreza Abdolrashidi and Lakshmish Ramaswamy

Session 2: Best Paper and Big Data for Enterprise

Evaluation and Analysis of In-Memory Key-Value Systems ......................................................................26
  Wenqi Cao, Semih Sahin, Ling Liu, and Xianqiang Bao

Improving the Visualization of WordNet Large Lexical Database through Semantic Tag Clouds .................................................................................................................................34
  Enrico Giacinto Caldarola and Antonio M. Rinaldi

Clustering Geo-tagged Tweets for Advanced Big Data Analytics ............................................................42
  Gloria Bordogna, Luca Frigerio, Alfredo Cuzzocrea, and Giuseppe Psaila
Session 3: Big Data Models and Algorithms

A NoSQL Data Model for Scalable Big Data Workflow Execution .......................................................52
   Aravind Mohan, Mahdi Ebrahimi, Shiyong Lu, and Alexander Kotov

Model Transformation and Data Migration from Relational Database to MongoDB ...........................................60
   Tianyu Jia, Xiaomeng Zhao, Zheng Wang, Dahan Gong, and Guiguang Ding

Predictive Modeling in a Big Data Distributed Setting: A Scalable Bias Correction Approach ........................................68
   Gianluca Bontempi and Yann-Aël Le Borgne

Session 4: Big Data Processing

Distributed Incremental Graph Analysis .................................................................................................75
   Upa Gupta and Leonidas Fegaras

Enhanced State History Tree (eSHT): A Stateful Data Structure for Analysis of Highly Parallel System Traces ..................................................................................................................83
   Loïc Prieur-Drevon, Raphaël Beamonte, Naser Ezzati-Jivan, and Michel R. Dagenais

A Case Study of Optimizing Big Data Analytical Stacks Using Structured Data Shuffling ........................................91
   Dixin Tang, Taoying Liu, Rubao Lee, Hong Liu, and Wei Li

Session 5: Big Data Analytics and Metrics

Geelytics: Enabling On-Demand Edge Analytics over Scoped Data Sources ...........................................101
   Bin Cheng, Apostolos Papageorgiou, and Martin Bauer

Analytics Toolkit for Business Big Data ..................................................................................................109
   Fan Liang and Weichang Du

QDrill: Query-Based Distributed Consumable Analytics for Big Data ..................................................117
   Shadi Khalifa, Patrick Martin, Dan Rope, Mike McRoberts, and Craig Statchuk

Session 6: Big Data Applications

Marketing Campaigns in Twitter Using a Pattern Based Diffusion Policy ............................................125
   Eleanna Kafeza, Christos Makris, and Pantelis Vikatos

TwigStack-MR: An Approach to Distributed XML Twig Query Using MapReduce ....................................133
   Hongjie Fan, Han Yang, Zhiyi Ma, and Junfei Liu

Infra: SLO Aware Elastic Auto-scaling in the Cloud for Cost Reduction .............................................141
   Subhajit Sidhanta and Supratik Mukhopadhyay
Session 7: Big Data Management

GOMA: Supporting Big Data Analytics with a Goal-Oriented Approach ...........................................................149
   Sam Supakkul, Liping Zhao, and Lawrence Chung

Energy Consumption Prediction with Big Data: Balancing Prediction Accuracy
and Computational Resources ..........................................................................................................................157
   Katarina Grolinger, Miriam A.M. Capretz, and Luke Seewald

From Big Data to Great Services ................................................................................................................165
   Jianwei Yin, Yan Tang, Wei Lo, and Zhaohui Wu

Session 8: Big Data Privacy

A Fast and Efficient Entity Resolution Approach for Preserving Privacy in Mobile Data .................................................................173
   Ioannis Boutsis and Vana Kalogeraki

SD-HDFS: Secure Deletion in Hadoop Distributed File System ..................................................................................181
   Bikash Agrawal, Raymond Hansen, Chunming Rong, and Tomasz Wiktorski

Privacy Preserving Predictive Analytics with Smart Meters .........................................................................................190
   B. Habtemariam, A. Miranskyy, A. Miri, S. Samet, and M. Davison

Session 9: Big Data Search and Mining

Catching Social Butterflies: Identifying Influential Users of an Event-Based Social Networking Service .................................................................198
   Jonathan Popa, Kusha Nezafati, Yulia R. Gel, John Zweck, and Georgiy Bobashev

Towards an Efficient Top-K Trajectory Similarity Query Processing Algorithm for Big Trajectory Data on GPGPUs ..................................................................................206
   Eleazar Leal, Le Gruenwald, Jianting Zhang, and Simin You

   Xiaoyong Yuan, Min Li, Sudeep Gaddam, Xiaolin Li, Yinan Zhao, Jingzhe Ma, and Jian Ge

BigData Congress Application Track

Session 1: Big Data Applications

Green Cabs vs. Uber in New York City .....................................................................................................................222
   Lasse Korsholm Poulsen, Daan Dekkers, Nicolaas Wagenaar, Wesley Snijders, Ben Lewinsky, Raghava Rao Mukkamala, and Ravi Vatrapu

Driver Unique Acceleration Behaviours and Stability over Two Years ........................................................................230
   Bruce Wallace, Rafik Goubran, Frank Knoefel, Shawn Marshall, Michelle Porter, and Andrew Smith
Session 2: Text Processing

Cloud-Based Core Text Processing Services for Sentiment Analysis .........................................................236

Huan Chen, Xin-Nan Li, Liang-Jie Zhang, Yi-Xuan Huang, and Xiao-Sheng Cai

Design and Implementation of a Multidimensional Data Retrieval Sorting Optimization Model ..........................................................244

Danfeng Yan, Liying Zhang, and Xuan Zhao

An Improved Single-Pass Algorithm for Chinese Microblog Topic Detection and Tracking ..........................................................251

Danfeng Yan, Enzheng Hua, and Bo Hu

Session 3: Big Data Streaming

Edge-Computing-Aware Deployment of Stream Processing Tasks Based on Topology-External Information: Model, Algorithms, and a Storm-Based Prototype ..................................................................................................................259

Apostolos Papageorgiou, Ehsan Poormohammady, and Bin Cheng

Effective Multi-stream Joining in Apache Samza Framework ..................................................................................267

Zhenyun Zhuang, Tao Feng, Yi Pan, Haricharan Ramachandra, and Badri Sridharan

Session 4: Evaluation

A Scalable Probabilistic Change Detection Algorithm for Very High Resolution (VHR) Satellite Imagery ............................................................................................................................275

Seokyong Hong and Ranga Raju Vatsavai

Evaluation of Data Mining Tools for Telecommunication Monitoring Data Using Design of Experiment ...............................................................................................................................283

Samneet Singh, Yan Liu, Wayne Ding, and Zheng Li


Man Si, Chung-Horng Lung, Samuel Ajila, and Wayne Ding

Session 5: Big Data Privacy, Security, and Economics

Genetic Information Privacy in the Age of Data-Driven Medicine .............................................................................299

Jingquan Li

A Method to Price Your Information Asset in the Information Market .............................................................................307

Divya Rao and Ng Wee Keong
BigData Congress Short Paper Track

Session 1: Big Data Processing
Workflow Transformation for Real-Time Big Data Processing .......................................................... 315
Yuji Ishizuka, Wuhui Chen, and Incheon Paik
An OWL Ontology Representation for Machine-Learned Functions Using Linked Data ................................. 319
Jingyuan Xu, Hao Wang, and Henry Trimbach
Quantitative Analysis of Scalable NoSQL Databases ............................................................................. 323
Surya Narayanan Swaminathan and Ramez Elmasri

Session 2: Big Data Management
Asset-centric Security-Aware Service Selection ....................................................................................... 327
Giannis Tziakouris, Marios Zinonos, Tom Chothia, and Rami Bahsoon
Semi-clustering That Scales: An Empirical Evaluation of GraphX ........................................................... 333
Jakob Smedegaard Andersen and Olaf Zukunft
Monetizing the User's Information Asset in Internet Information Market .................................................. 337
Divya Rao and Wee Keong Ng

Session 3: Big Data Systems and Architecture
Privacy-Aware Big Data Warehouse Architecture ....................................................................................... 341
Karthik Navuluri, Ravi Mukkamala, and Aftab Ahmad
Software Metrics for Green Parallel Computing of Big Data Systems .................................................... 345
Havva Gölay Gürbüz and Bedir Tekinerdogan

Session 4: Big Data Systems and Mining
Distributed Top-k Keyword Search over Very Large Databases with MapReduce ........................................ 349
Ziqiang Yu, Xiaohui Yu, Yuehui Chen, and Kun Ma
Towards Automatic Memory Tuning for In-Memory Big Data Analytics in Clusters ................................. 353
Aris-Kyriakos Koliopoulos, Paraskevas Yiapanis, Firat Tekiner, Goran Nenadic, and John Keane
Analyzing Future Nodes in a Knowledge Network ..................................................................................... 357
Sukhwam Jung, Tuan Manh Lai, and Aviv Segev
BigData Congress Special Track

Session 1: Big Data Applications
Do Drivers' Behaviors Reflect Their Past Driving Histories? - Large Scale Examination of Vehicle Recorder Data
Daisaku Yokoyama and Masashi Toyoda

Urban Area Vehicle Number Estimation Based on RTMS Data
Yue Hu, Yuanchao Shu, Peng Cheng, and Jiming Chen

Towards Data Analytics of Pathogen-Host Protein-Protein Interaction: A Survey
Huaming Chen, Jun Shen, Lei Wang, and Jiangning Song

Session 2: Big Data Management
Open Source Big Data Analytics Frameworks Written in Scala
John A. Miller, Casey Bowman, Vishnu Gowda Harish, and Shannon Quinn

Big Datasets for Research: A Survey on Flagship Conferences
Yi Wei, Shijun Liu, Jiao Sun, Lizhen Cui, Li Pan, and Lei Wu

Mobile Network Traffic Prediction Using MLP, MLPWD, and SVM
Ali Yadavar Nikravesh, Samuel A. Ajila, Chung-Horng Lung, and Wayne Ding

Session 3: Evaluation and Operations
JVM Configuration Management and Its Performance Impact for Big Data Applications
Semih Sahin, Wenqi Cao, Qi Zhang, and Ling Liu

An Hybrid Approach to Quality Evaluation across Big Data Value Chain
Mohamed Adel Serhani, Hadeel T. El Kassabi, Ikbal Taleb, and Alramzana Nujum

BigData Congress Services Special Track

Session 1: Service Applications
GS1 Connected Car Using EPCIS-ONS System
Bongjin Sohn, Sungpil Woo, Jiyong Han, Hyeeun Cho, Jaewook Byun, and Daeyoung Kim

Key Update as a Service (KAAS): An Agent-Based Modeling for Cloud-Based Access Control
Somchart Fugkeaw and Hiroyuki Sato
Session 2: Service Modeling

Modeling the Location Selection of Mirror Servers in Content Delivery Networks .................................................................................................................................438

Peter Hillmann, Tobias Uhlig, Gabi Dreo Rodosek, and Oliver Rose

Cloud Computing Intelligent Data-Driven Model: Connecting the Dots to Combat Global Terrorism .................................................................................................................................446

Gokop L. Goteng and Xueyu Tao

Research of Clouding Based Pulse Monitoring and Data Analysis Framework for Efficient Physical Training ........................................................................................................................................454

Hong Liang Yuan, Jun Wang, Jun Liu, and Shi Liang Li

Session 3: QoS Management

Complex Quality of Service Lifecycle Assessment Methodology .................................................................................................................................462

R. William Maule

Don't Fire Me, a Kernel Autoregressive Hybrid Model for Optimal Layoff Plan .................................................................................................................................470

Zhiling Luo, Ying Li, Ruisheng Fu, and Jianwei Yin

Can We Use Programmer's Knowledge? Fixing Parameter Configuration Errors in Hadoop through Analyzing Q&A Sites ........................................................................................................................................478

Jia Tong, Li Ying, Tang Hongyan, and Wu Zhonghai

Session 4: Services Clustering and Discovery

Identification as a Service: Large-Scale Cloud Service Discovery over the World Wide Web .................................................................................................................................485

Abdullah Alfazi, Quan Z. Sheng, Wei Emma Zhang, Lina Yao, and Talal H. Noor

Software Evolution Information Driven Service-Oriented Software Clustering .................................................................................................................................493

Linhui Zhong, Jing He, Nengwei Zhang, Peng Zhang, and Jing Xia

Author Index ........................................................................................................................................501