

2017 Fifth International Conference on Advanced Cloud and Big Data (CBD 2017)

**Shanghai, China
13-16 August 2017**



**IEEE Catalog Number: CFP1703Y-POD
ISBN: 978-1-5386-1073-2**

**Copyright © 2017 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

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

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

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

IEEE Catalog Number:	CFP1703Y-POD
ISBN (Print-On-Demand):	978-1-5386-1073-2
ISBN (Online):	978-1-5386-1072-5

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

2017 Fifth International Conference on Advanced Cloud and Big Data

CBD 2017

Table of Contents

Welcome Message	xi
Organizing Committee	xii
Program Committee	xiii
Reviewers	xvi

Session 1: Keynote

Business Intelligence Revisited	1
<i>Jianwen Su and Yan Tang</i>	

Session 2: Cloud Computing and Applications I

A Bandwidth Allocation Policy for Helpers in Cloud-Assisted P2P Video-on-Demand Systems	7
<i>Guowei Huang, Lingjing Kong, Keke Wu, and Zhi Chen</i>	
A Dual-Buffer Based Acceleration Strategy for FW Block Algorithm on GPUS	13
<i>Yiwen Li and Jian Yu</i>	
A Large Scale Data Transmission Control Mechanism across Data Centers	20
<i>Wenjiang Ding, Wei Li, Runhuan Zhang, Qianqian Zhang, and Chenhao Yang</i>	
A Novel Solution of Cloud Operating System Based on X11 and Docker	26
<i>Zhen Du, Zhuging Xu, Fang Dong, and Dian Shen</i>	
Adaptive Resource Prediction in the Cloud Using Linear Stacking Model	33
<i>Shasha Liao, Hongjie Zhang, Guansheng Shu, and Jing Li</i>	
An Heuristic for Bag-of-Tasks Scheduling Problems with Resource Demands and Budget Constraints to Minimize Makespan on Hybrid Clouds	39
<i>Yi Zhang, Jin Sun, and Zebin Wu</i>	
BigVM: A Multi-Layer-Microservice-Based Platform for Deploying SaaS	45
<i>Tianlei Zheng, Yuqun Zhang, Xi Zheng, Min Fu, and Xiao Liu</i>	

Cost Optimization for Scheduling Scientific Workflows on Clouds under Deadline Constraints	51
<i>Wei Zheng, Bugingo Emmanuel, Chen Wang, Yingsheng Qin, and Dongzhan Zhang</i>	

Session 3: Cloud Computing and Applications II

Dynamic Flow Scheduling for Power Optimization of Data Center Networks	57
<i>Juan Luo, Song Zhang, Luxiu Yin, and Yaling Guo</i>	
Efficient Data and Task Co-Scheduling for Scientific Workflow in Geo-Distributed Datacenters	63
<i>Jian Chen, Jinghui Zhang, and Aibo Song</i>	
EQF: An Explicit Queue-Length Feedback for TCP Congestion Control in Datacenter Networks	69
<i>Yifei Lu, Xiaoting Fan, and Lei Qian</i>	
Estimating SDN Traffic Matrix Based on Online Informative Flow Measurement Method	75
<i>Dongyang Li, Ningyun Dai, Feng Li, Changyou Xing, and Fei Dai</i>	
PCMSim: A Hybrid Memory System Simulator for the Cloud Storage	81
<i>Jibin Wang and Bohan Wang</i>	
Research and Implementation of a Software Online Testing Platform Model Based on Cloud Computing	87
<i>Jing Chen, Chunxiao Wang, Fei Liu, and Yinglong Wang</i>	
Towards Cost Reduction in Cloud-Based Workflow Management through Data Replication	94
<i>Fei Xie, Jun Yan, and Jun Shen</i>	

Session 4: Big Data and Applications I

A Cost-Efficient Data Placement Algorithm with High Reliability in Hadoop	100
<i>Yao Du, Runqun Xiong, Jiahui Jin, and Junzhou Luo</i>	
A Data Replenishment Method for Self Regulating Sail-Assisted Ships Considering Big Data	106
<i>Tianyi Zhang, Jianing Zhang, and Chengming Yang</i>	
A Decision Support Scheme for Cloud Storage	111
<i>Yang Gao, Keqiu Li, and Heng Qi</i>	
A Distributed Approach for Constructing Generalized Suffix Tree on Spark by Using Optimized Elastic Range Algorithm	117
<i>Yunhao Li, Jiahui Jin, Runqun Xiong, and Junzhou Luo</i>	
A Feature-Based Food Recommendation on Apache Spark	123
<i>Wenyuan Xu, Yun Li, Liting Wei, and Zheng Wu</i>	

An Online Prediction Algorithm of Traffic in Big Data Based on the Storm	129
<i>Shangdong Liu, Yimu Ji, Dianchao Zhang, Yongge Yuan, Jian Gong, and Ruchuan Wang</i>	
An Ontology-Based Data Organization Method	135
<i>Qian Hao, Yue Li, Li Min Wang, and Mei Wang</i>	

Session 5: Big Data and Applications II

Consistent Hash Based Flexible Redundant Data Partition Strategy	141
<i>Zhiyun Zheng, Ning Wang, Dun Li, Zhimeng Du, and Zhenfei Wang</i>	
Data Deduplication Based on Hadoop	147
<i>Dongzhan Zhang, Chengfa Liao, Wenjing Yan, Ran Tao, and Wei Zheng</i>	
Detection of Super Nodes for Network Data Streams with Filtering Strategy	153
<i>Aiping Zhou, Huisheng Zhu, Lijun Liu, and Chengang Zhu</i>	
Distributed Processing Practice of the 3D City Model Based on HBase	159
<i>Chaoquan Zhang and Bo Mao</i>	
Finding Compositional Skyline Based on Sharing Strategy	164
<i>Leigang Dong and Guohua Liu</i>	
GeaBase: A High-Performance Distributed Graph Database for Industry-Scale Applications	170
<i>Zhisong Fu, Zhengwei Wu, Houyi Li, Yize Li, Min Wu, Xiaojie Chen, Xiaomeng Ye, Benquan Yu, and Xi Hu</i>	
Intelligent Anti-Money Laundering Solution Based upon Novel Community Detection in Massive Transaction Networks on Spark	176
<i>Xurui Li, Xiang Cao, Xuetao Qiu, Jintao Zhao, and Jianbin Zheng</i>	
Logic-Partition Based Gaussian Sampling for Online Aggregation	182
<i>Longbin Zhang, Yuxiang Wang, and Xiaoliang Xu</i>	

Session 6: Social Computing and Applications I

A Lexical Updating Algorithm for Sentiment Analysis on Chinese Movie Reviews	188
<i>Yiwei Song, Kaiwen Gu, Huakang Li, and Guozi Sun</i>	
A Multidimensional Comprehensive Recommendation Method Based on Social Network	194
<i>Bo Zhang, Cui Mu, Qin Zhao, Zhangjie Peng, Junhong Ding, and Bo Liu</i>	
A Novel Method for Open Relation Extraction from Public Announcements of Chinese Listed Companies	200
<i>Tianyuan Zhuang, Pengwei Wang, Yaying Zhang, and Zhaohui Zhang</i>	
A Real-Time Bus Transfer Scheme Recommendation Systems	206
<i>Jun Zhang, Lai Tu, Fan Zhang, and Benxiong Huang</i>	

A Search Strategy for Social Resource in Decentralized Social Networks	213
<i>Wenxiu Xu, Yonghong Guo, Leilei Shi, Lu Liu, and Bo Yuan</i>	
ABT: An Effective Ability-Balanced Team Based Incentive Mechanism in Crowdsourcing System	220
<i>Yufeng Wang, Jie Huang, Qun Jin, and Jianhua Ma</i>	
Community Discovery in Large-Scale Complex Networks Using Distributed SimRank Nonnegative Matrix Factorization	226
<i>Chaobo He, Xiang Fei, Hanchao Li, Hai Liu, Yong Tang, and Qimai Chen</i>	

Session 7: Social Computing and Applications II

Cost-Efficient Influence Maximization in Online Social Networks	232
<i>Jingya Zhou, Jianxi Fan, Jin Wang, Xi Wang, and Baolei Cheng</i>	
Efficient Fine-Grained Location Prediction Based on User Mobility Pattern in LBSNs	238
<i>Jiuxin Cao, Shuai Xu, Xuelin Zhu, Renjun Lv, and Bo Liu</i>	
K-Truss Based Top-Communities Search in Large Graphs	244
<i>Anzuruni Maulidi Katunka, Cairong Yan, Kossonon Brandou Serge, and Zhaohui Zhang</i>	
Location and Motion Prediction of Consumers in a Large Shopping Mall	250
<i>Zeyu Zhang and Weiping Zhu</i>	
Passenger Travel Path Estimation Algorithm Based on High Accuracy Location Data	256
<i>Huali Xiao, Li Sun, Shaohong Kong, Guiwei Gong, and Fan Zhang</i>	
VOut: Mining Sampled GPS Trace for Taximeter Fraud	261
<i>Ding Jing and Hongzi Zhu</i>	

Session 8: Cloud Security and Applications

Dynamic Cloud Access Control via Accelerometers Using Customized Actions of Occupants	267
<i>Biaokai Zhu, Jumin Zhao, Deng-Ao Li, Hong Wang, Ruiqin Bai, Yanxia Li, and Bin Yang</i>	
Fast Identification of Encrypted Traffic via Large-Scale Sparse Screening	273
<i>Peng Meng, Guopeng Zhou, and Juan Meng</i>	
Identifying Mobile Applications for Encrypted Network Traffic	279
<i>Gaofeng He, Bingfeng Xu, and Haiting Zhu</i>	
Detecting Clocking Attack in Contactless Access Control Systems	285
<i>Guixiang Tian, Li Lu, and Muhammad Jawad Hussain</i>	

Session 9: Machine Learning and Miscellaneous I

A Data Reconstruction Algorithm Based on Neural Network for Compressed Sensing	291
<i>Li Tian, Guorui Li, and Cong Wang</i>	
A Mining Approach to Obtain the Software Vulnerability Characteristics	296
<i>Xiang Li, Jinfu Chen, Zhechao Lin, Lin Zhang, Zibin Wang, Minmin Zhou, and Wanggen Xie</i>	
A New Design of Real-Time Traffic Index Model for Freeway	302
<i>Fusheng Xu, Zhongxiang Huang, and Xueying Zhu</i>	
A Parallel Algorithm for Mining Time Relaxed Gradual Clustering Pattern Based on Spatio-Temporal Trajectories	308
<i>Hongyan Sun, Genlin Ji, Bin Zhao, and Xintao Liu</i>	
An Improved Algorithm for Frequent Itemsets Mining	314
<i>Hao Jiang and Xu He</i>	
Deep Neural Network Based on Translation Model for Diabetes Knowledge Graph	318
<i>Suna Yin, Dehua Chen, and Jiajin Le</i>	
Detecting Spammers in E-Commerce Website via Spectrum Features of User Relation Graph	324
<i>Bo Guo, Hao Wang, Zhaojun Yu, and Yu Sun</i>	
Improved Candidate Answers Ranking for QA	331
<i>Chengchao Yu, Rui Zhao, and Zhihua Wei</i>	

Session 10: Machine Learning and Miscellaneous II

PixVid: Capturing Temporal Correlated Changes in Time Series	337
<i>Yeqiang Lin, Mehrzad Lavassani, Jiyun Li, and Tingting Zhang</i>	
RBTA: A Multivariate Time-Series Method for City Incidents Mining and Forecasting	343
<i>Jieyi Wang, Yongkun Wang, and Yaohui Jin</i>	
Recommendation for Cross-Disciplinary Collaboration Based on Potential Research Field Discovery	349
<i>Wei Liang, Xiaokang Zhou, Suzhen Huang, Chunhua Hu, and Qun Jin</i>	
Recommendation Algorithm Based on User Trust and Interest with Probability Matrix Factorization	N/A
<i>Chuanmin Mi, Peng Peng, Lin Xiao, and Yue Lu</i>	
Search for Top-K Spatial Objects Based on Various-Widths Clustering	362
<i>Shoujian Yu and Guangyi Feng</i>	

Semi-SGD: Semi-Supervised Learning Based Spammer Group Detection in Product Reviews	368
<i>Lu Zhang, Yang Yuan, Zhiang Wu, and Jie Cao</i>	
Target Distribution Guided Network Sampling	374
<i>Renjie Fan, Zhiwen Yu, Bin Guo, Liang Wang, and Dingqi Yang</i>	
Trajectory Data Driven Transit-Transportation Planning	380
<i>Yihan Guo, Shaoyong Wang, Lin Zheng, and Mingming Lu</i>	

Session 11: Machine Learning and Miscellaneous III

Two-Phase Neural Network Model for Pollution Concentrations Forecasting	385
<i>Cairong Yan, Shuhua Xu, Yan Huang, Yongfeng Huang, and Zhaohui Zhang</i>	
Predicting Object Types in Linked Data by Text Classification	391
<i>Xiang Zhang, Erjing Lin, and Siyao Pi</i>	
Video Object Detection for Tractability with Deep Learning Method	397
<i>Bing Tian, Liang Li, Yansheng Qu, and Li Yan</i>	
Your Body Defines Your Fall Detection System: A Somatotype-Based Feature Selection Method	402
<i>Xiaoling Fu, Zheng Tan, Chungang Yan, Zhong Li, Cheng Wang, and Cheng-Xiang Wang</i>	
An Indoor TV White Space Resource Collecting and APs Locations Selecting Scheme	408
<i>Chengsi Qi, Guanglin Zhang, Wenqiang Zhang, and Demin Li</i>	
An Outlier Detection Method and Its Application for Earthen Ruins Data Monitored by WSN	414
<i>Yun Xiao, Haohan Gao, Jia Ding, Xuanhong Wang, and Xianjia Meng</i>	
ORSCA: Optimal Route Selection and Communication Association for Drones in WSNs	420
<i>Tao Wu, Panlong Yang, Yubo Yan, Xunpeng Rao, Ping Li, and Wanru Xu</i>	
Author Index	425