

**2016 IEEE/ACM 2nd  
International Workshop on Big  
Data Software Engineering  
(BIGDSE 2016)**

**Austin, Texas, USA  
16 May 2016**



**IEEE Catalog Number: CFP16C45-POD  
ISBN: 978-1-5090-2195-6**

**Copyright © 2016, Association for Computing Machinery  
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:	CFP16C45-POD
ISBN (Print-On-Demand):	978-1-5090-2195-6
ISBN (Online):	978-1-4503-4152-3

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# 2nd International Workshop on BIG Data Software Engineering (BIGDSE 2016)

## Table of Contents

Message from Workshop Chairs .....	vii
Committees.....	viii
Sponsors and Supporters .....	x

### Session 1: Real-Time Data Analytics

Decisions as a Service for Application Centric Real Time Analytics .....	1
<i>Patrick Tendick and Audris Mockus</i> —Avaya Labs Research, USA	
Streaming Software Analytics .....	8
<i>Georgios Gousios, Dominik Safaric, and Joost Visser</i> —Radboud University Nijmegen, Netherlands	

### Session 2: Data Management

DataLab: A Version Data Management and Analytics System.....	12
<i>Yang Zhang, Fangzhou Xu, Erwin Frise, Siqi Wu, Bin Yu, and Wei Xu</i> —Tsinghua University, China; UC Berkeley, USA	
Exploring a Framework for Identity and Attribute Linking across Heterogeneous Data Systems .....	19
<i>Nathan Wilder, Jared M. Smith, and Audris Mockus</i> —University of Tennessee, USA	
Providing Big Data Applications with Fault-Tolerant Data Migration across Heterogeneous NoSQL Databases .....	26
<i>Marco Scavuzzo, Damian A. Tamburri, and Elisabetta di Nitto</i> —Politecnico di Milano - DEIB, Italy	
Data Model Evolution Using Object-NoSQL Mappers: Folklore or State-of-the-Art? .....	33
<i>Andreas Ringlstetter, Stefanie Scherzinger, and Tegawendé F. Bissyande</i> —OTH Regensburg, Germany; University of Luxembourg, Luxembourg	

## Session 3: Engineering Data-Intensive Systems

Towards a Model-Driven Design Tool for Big Data Architectures .....	37
<i>Michele Guerriero, Saeed Tajfar, Damian Andrew Tamburri, and Elisabetta di Nitto</i> —Politecnico di Milano, Italy	
Toward Big Data Value Engineering for Innovation.....	44
<i>Hong-Mei Chen, Rick Kazman, Juan Garbajosa, and Eloy Gonzalez</i> —University of Hawaii at Manoa, USA; Software Engineering Institute, USA; Technical University of Madrid (UPM), Spain; Indra Software Labs, Spain	
Understanding Quality Requirements in the Context of Big Data Systems.....	76
<i>Ibtehal Noorwali, Darlan Arruda, and Nazim H. Madhavji</i> —University of Western Ontario, Canada	

## Session 4: Applications and Industry

A Reference Architecture for Big Data Systems in the National Security Domain .....	51
<i>John Klein, Ross Buglak, David Blockow, Troy Wuttke, and Brenton Cooper</i> —Software Engineering Institute, USA; D2D CRC, Australia	
A Big Data Framework for Cloud Monitoring .....	58
<i>Saeed Zareian, Marios Fokaefs, Hamzeh Khazaei, Marin Litoiu, and Xi Zhang</i> —York University, Canada; Sichuan University, China	
The 'BigSE' Project: Lessons Learned from Validating Industrial Text Mining .....	65
<i>Rahul Krishna, Zhe Yu, Amritanshu Agrawal, Manuel Dominguez, and David Wolf</i> —North Carolina State University, USA; LexisNexis, USA	
Predicting and Fixing Vulnerabilities before They Occur: A Big Data Approach .....	72
<i>Hong-Mei Chen, Rick Kazman, Ira Monarch, and Ping Wang</i> —University of Hawaii, USA; Software Engineering Institute, USA; Carnegie-Mellon University, USA; IPTI, USA; University of Maryland, USA	
<b>Author Index .....</b>	<b>80</b>