

# SOFTWARE ARCHITECTURE FOR BIG DATA AND THE CLOUD

EDITED BY IVAN MISTRIK, RAMI BAHSOON, NOUR ALI,  
MARITTA HEISEL, BRUCE MAXIM

**MK**  
MORGAN KAUFMANN



# SOFTWARE ARCHITECTURE FOR BIG DATA AND THE CLOUD

Applied cloud computing for data intensive science and software architecture

Edited by:

Ivan Mistrik, Rami Bahsoon, Nour Ali, Maritta Heisel, Bruce Maxim

With Forewords from:

- Mandy Chessell, IBM's Hursley Laboratory, UK, writing on the movement to Big Data, Amnesia or progress?
- Ian Gorton, North Eastern University, Seattle, Washington, USA, on focusing the spotlight of software architecture research on scalability

*Software Architecture for Big Data and the Cloud* is designed to be a single resource that brings together research on how software architectures can solve the challenges imposed by building big data software systems. The challenges of big data on the software architecture can relate to scale, security, integrity, performance, concurrency, parallelism, dependability, among many others. Big data handling requires rethinking architectural solutions to meet functional and non-functional requirements related to volume, variety and velocity, their development and evolution trends in relation to the operating environment. With the wide adoption of the cloud as an enabler environment for exploiting data potentials as service under new economics models, little is known about the underlying architectures which can sustain this synergy. The editors have varied and complementary backgrounds in requirements and architecture, specifically in software architectures for cloud and big data, as well as expertise in software engineering for cloud and big data. This book collects together work across different disciplines in software engineering for cloud and big data, including work expanded from conference tracks and workshops led by the editors.

Key Features:

- Discusses systematic and disciplined approaches to building software architectures for cloud and big data with state-of-the-art methods and techniques
- Presents case studies involving enterprise, business, and government service deployment of big data applications
- Shares guidance on theory, frameworks, methodologies, and architecture for cloud and big data

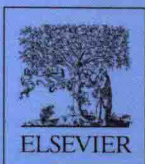
Ivan Mistrik, Independent researcher in software-intensive systems engineering, Germany

Rami Bahsoon, Senior lecturer in Software Engineering and founder of the Cloud Software Engineering interest group at the School of Computer Science, the University of Birmingham, UK

Nour Ali, Senior lecturer at the University of Brighton, UK

Maritta Heisel, Professor for Software Engineering at the University Duisburg-Essen, Germany

Bruce Maxim, Associate professor of Computer and Information Science at the University of Michigan—Dearborn, MI, USA



# MK

MORGAN KAUFMANN PUBLISHERS

AN IMPRINT OF ELSEVIER

[elsevier.com/books-and-journals](http://elsevier.com/books-and-journals)

Computers/  
Software Development  
and Engineering

ISBN 978-0-12-805467-3



9 780128 054673

# SOFTWARE ARCHITECTURE FOR BIG DATA AND THE CLOUD

EDITED BY  
MISTRIK,  
BAHsoon,  
ALI,  
HEISEL,  
MAXIM

MK

# Software Architecture for Big Data and the Cloud

Edited by

**Ivan Mistrik**

**Rami Bahsoon**

**Nour Ali**

**Maritta Heisel**

**Bruce Maxim**



**MK**

MORGAN KAUFMANN PUBLISHERS

AN IMPRINT OF ELSEVIER

Morgan Kaufmann is an imprint of Elsevier  
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

Copyright © 2017 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: [www.elsevier.com/permissions](http://www.elsevier.com/permissions).

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

## Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

## Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

## British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-0-12-805467-3

For information on all Morgan Kaufmann publications  
visit our website at <https://www.elsevier.com/books-and-journals>



Working together  
to grow libraries in  
developing countries

[www.elsevier.com](http://www.elsevier.com) • [www.bookaid.org](http://www.bookaid.org)

*Publisher:* Todd Green, Jonathan Simpson

*Acquisition Editor:* Todd Green, Jonathan Simpson

*Editorial Project Manager:* Lindsay Lawrence

*Senior Production Project Manager:* Priya Kumaraguruparan

*Designer:* Vicky Pearson Esser

Typeset by VTeX

# Software Architecture for Big Data and the Cloud





# CONTRIBUTORS

**Nour Ali**

University of Brighton, UK

**Katerina Argyraki**

EPFL, Switzerland

**Ioannis N. Athanasiadis**

Wageningen University, Information Technology, Wageningen, The Netherlands

**Cigdem Avci Salma**

Wageningen University, Information Technology, Wageningen, The Netherlands

**Rami Bahsoon**

School of Computer Science, University of Birmingham, UK

**Alistair Barros**

Science and Engineering Faculty, Queensland University of Technology (QUT), Brisbane, QLD, Australia

**Stephen Bonner**

Institute of Advanced Research Computing, Durham University, United Kingdom

**John Brennan**

Institute of Advanced Research Computing, Durham University, United Kingdom

**Rajkumar Buyya**

Cloud Computing and Distributed Systems (CLOUDS) Laboratory, School of Computing and Information Systems, The University of Melbourne, Australia

**Bo Chen**

Department of Computer Science, University of Memphis, Memphis, TN, USA

**Tao Chen**

School of Computer Science, University of Birmingham, UK

**Mandy Chessell**

IBM, Winchester, Hampshire, United Kingdom

**Jose G.F. Coutinho**

Imperial College London, UK

**Reza Curtmola**

Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, USA



**Jun Dai**

Southwestern University of Finance and Economics, Chengdu, Sichuan, China  
Rutgers University–Newark, Newark, NJ, USA

**Anastasija Efremovska**

Vrije Universiteit Amsterdam, The Netherlands

**Robert Eikermann**

Software Engineering, RWTH Aachen University, Aachen, Germany

**Colin Fidge**

Science and Engineering Faculty, Queensland University of Technology (QUT), Brisbane, QLD, Australia

**Andrei Furda**

Science and Engineering Faculty, Queensland University of Technology (QUT), Brisbane, QLD, Australia

**Saurabh Garg**

School of Engineering and ICT, University of Tasmania, Tasmania, Australia

**Ian Gorton**

Northeastern University, United States

**Wilhelm Hasselbring**

Kiel University, Germany

**Robert Heinrich**

Karlsruhe Institute of Technology, Germany

**Maritta Heisel**

Universität Duisburg-Essen, Germany

**George Ioannidis**

EPFL, Switzerland

**Anca Iordache**

Université de Rennes 1, France

**Reiner Jung**

Kiel University, Germany

**Christoph Kleineweber**

Zuse Institute Berlin, Germany

**Alexandros Koliouris**

Imperial College London, UK

**Ibad Kureshi**

Institute of Advanced Research Computing, Durham University, United Kingdom

**Patricia Lago**

Vrije Universiteit Amsterdam, The Netherlands

**Markus Look**

Software Engineering, RWTH Aachen University, Aachen, Germany

**Darshan Lopes**

Capgemini UK, UK

**Bruce Maxim**

University of Michigan-Dearborn, USA

**John McGlone**

SAP Labs, USA

**Ivan Mistrik**

Independent Researcher, Germany

**Alp Oral**

Wageningen University, Information Technology, Wageningen, The Netherlands

Microsoft, Vancouver, Canada

**Fiona O'Sullivan**

Capgemini UK, UK

**Kevin Palmer**

Capgemini UK, UK

**Guillaume Pierre**

Université de Rennes 1, France

**Deepak Poola**

The University of Melbourne, Australia

**Carlos Queiroz**

The Association of Computing Machinery (ACM), Singapore

**Carmelo Ragusa**

SAP, UK

**Kotagiri Ramamohanarao**

The University of Melbourne, Australia

**Ralf Reussner**

Karlsruhe Institute of Technology, Germany

**Maria A. Rodriguez**

Cloud Computing and Distributed Systems (CLOUDS) Laboratory, School of Computing and Information Systems, The University of Melbourne, Australia

**Alexander Roth**

Software Engineering, RWTH Aachen University, Aachen, Germany

**Berhard Rumpe**

Software Engineering, RWTH Aachen University, Aachen, Germany

**Mark Ryan**

School of Computer Science, University of Birmingham, UK

**Mohsen Amini Salehi**

The University of Louisiana Lafayette, USA

**Peter Sanders**

Maxeler Technologies, UK

**Thorsten Schütt**

Zuse Institute Berlin, Germany

**Mark Stillwell**

Imperial College London, UK

**Bedir Tekinerdogan**

Wageningen University, Information Technology, Wageningen, The Netherlands

**Georgios Theodoropoulos**

Institute of Advanced Research Computing, Durham University, United Kingdom

**Tim Vincent**

IBM, Toronto, Ontario, Canada

**Alexander Wolf**

University of California, Santa Cruz, USA

**Dan Wolfson**

IBM, Austin, TX, USA

**Andreas Wortmann**

Software Engineering, RWTH Aachen University, Aachen, Germany

**Jiangshan Yu**

School of Computer Science, University of Birmingham, UK

**Teng Yu**

Imperial College London, UK

**Xinwei Zhao**

School of Engineering and ICT, University of Tasmania, Tasmania, Australia

**Olaf Zimmermann**

University of Applied Sciences of Eastern Switzerland (HSR FHO), Rapperswil, Switzerland

**Christian Zirkelbach**

Kiel University, Germany





# ABOUT THE EDITORS

**Ivan Mistrik** is a researcher in software-intensive systems engineering. He is a computer scientist who is interested in system and software engineering and in system and software architecture, in particular: life cycle system/software engineering, requirements engineering, relating software requirements and architectures, knowledge management in software development, rationale-based software development, aligning enterprise/system/software architectures, value-based software engineering, agile software architectures, and collaborative system/software engineering. He has more than 40 years' experience in the field of computer systems engineering as an information systems developer, R&D leader, SE/SA research analyst, educator in computer sciences, and ICT management consultant. During the past 40 years, he has been primarily working at various R&D institutions in USA and Germany and has done consulting on a variety of large international projects sponsored by ESA, EU, NASA, NATO, and UN. He has also taught university-level computer sciences courses in software engineering, software architecture, distributed information systems, and human–computer interaction. He is the author or coauthor of more than 90 articles and papers in international journals, conferences, books, and workshops. Ivan Mistrik has written a number of editorials for special issues and edited books. He has also written over 120 technical reports and presented over 70 scientific/technical talks. He has served in many program committees and panels of reputable international conferences and organized a number of scientific workshops. He was the lead-editor of ten books between 2006 and 2016: *Rationale Management in Software Engineering*; *Rationale-Based Software Engineering*; *Collaborative Software Engineering*; *Relating Software Requirements and Architecture*; *Aligning Enterprise, System, and Software Architectures*; *Agile Software Architecture*; *Economics-driven Software Architecture*; *Relating System Quality and Software Architecture*; *Software Quality Assurance*; *Managing Trade-Offs in Adaptable Software Architectures*.

**Nour Ali** is a Principal Lecturer at the University of Brighton since December, 2012. She holds a PhD in Software Engineering from the Polytechnic University of Valencia-Spain for her work in Ambients in Aspect-Oriented Software Architecture. She is a Fellow of UK Higher Education Academy (HEA). Her research area encompasses service oriented architecture, software architecture, self-adaptation and mobile systems. In 2014, the University of Brighton granted her a Rising Stars award in Service Oriented Architecture Recovery and Consistency. She is currently the Principal Investigator for the Royal Society Newton grant, “An Autonomic Architectural Approach for Health Care Systems,” and is the Knowledge Supervisor for the Knowledge Transfer Partnership project for migrating legacy software systems using architecture centric approach. She has also been the Principal Investigator for an Enterprise Ireland Commercialisation Project in Architecture Recovery and Consistency and coinvestigator in several funded projects. Dr. Ali serves on the Programme Committee for several conferences (e.g., ICWS, ICMS, HPCC, etc.) and journals (e.g., JSS and JIST). She has cochaired and coorganized several workshops such as the IEEE International Workshop on Engineering Mobile Service Oriented Systems (EMSOS) and the IEEE Workshop on Future of Software Engineering for/in the Cloud. She was the coeditor of the JSS Special Issue on the Future of Software Engineering for/in the Cloud published in 2013 and has coedited three books including “Agile and lean service-oriented development: foundations, theory, and practice” published in 2012. She is the Application

Track chair for International Conference on Web Services (ICWS 2016). Her personal website is: <http://www.cem.brighton.ac.uk/staff/na179/>.

**Rami Bahsoon** is a Senior Lecturer Software Engineering (Assoc. Prof.) and founding member of the Software Engineering Research group and the Cloud Software Engineering interest group at the University of Birmingham, UK. He holds a PhD in Software Architecture from University College London. During his PhD, he was awarded a fellowship to attend London Business School (LBS) for MBA studies in technology strategy and strategy dynamics. Bahsoon's research interests are in Cloud software Engineering, Software Architectures, Self-adaptive and Managed Software Engineering, Security Software Engineering, Relating software requirements (nonfunctional requirements) to software architectures, testing and regression testing, software maintenance and evolution, software metrics, empirical evaluation, Software Sustainability and Economics-driven Software Engineering research. His Birmingham Group comprises 10 PhD students and was among the first groups worldwide to carry fundamental research on cloud software engineering and software architectures evidenced by publications to top-tier venues such as IEEE Transactions on Cloud Computing, IEEE Transactions on Software Engineering, IEEE Transactions on Services Computing, IEEE Computer, ICWS, IEEE Cloud, WICSA, SEAMS/ICSE, UCC, HPCC, among others. His group has been working on self-adaptive and managed architectures for supporting the development and evolution of dependable ultra-large software systems covering cloud; dynamic resource allocation and federation in cloud; self-aware cloud; cloud elasticity, autoscaling and dynamic management of Quality of Service (QoS); utility models for service composition; volunteer services computing. Bahsoon had founded the IEEE International Software Engineering IN/FOR the Cloud workshop (in its 6th version now) in conjunction with IEEE Services and IEEE Cloud, the leading venue for cloud research. He was the lead editor for a special issue on the Future of Software Engineering for/In the Cloud and another on Architecture and Mobility and with the Journal of Systems and Software, a leading software engineering journal. Bahsoon has coedited a book on Software Architectures for Cloud and Big Data (Elsevier). He coedited another three books on Software Architecture and Software Quality (Elsevier 2014); Economics-Driven Software Architecture (Elsevier 2014); Aligning Enterprise, System, and Software Architectures (IGI Global in 2012). He has acted as the workshop chair for IEEE Services 2014, IEEE Cloud, IEEE Big Data, ICWS; the Doctoral Symposium chair of IEEE/ACM Utility and Cloud Computing Conference (UCC 2014 and UCC2016); track chair for Utility and Cloud Computing of IEEE HPCC 2014; Visionary Track chair for IEEE Services (2015); Emerging Technologies Track Chair for IEEE Services (2016); IEEE Big Data UK Satellite session chair (2015) on Big Data Software Engineering for Cloud and Mobile Services (2015); Big Data Software Engineering for Cloud, Edge Computing and Mobility; workshop chair for ECSA 2016 and programme chair for IEEE Services 2017. He is a member of EPSRC Associate College Board and acted as a panellist for the NSF/EPSRC Software Grand Challenge for expertise in cloud software engineering, had reviewed for major EU/UK funding bodies and software engineering and service science conferences and journals. His website is: <https://www.cs.bham.ac.uk/~rzb/>.

**Maritta Heisel** is a Full Professor for software engineering at the University Duisburg-Essen, Germany, since 2004. Her research interests include the development of dependable software, pattern- and component-based software development, requirements engineering (including quality requirements), software architecture, and software evolution. She is particularly interested in incorporating security and privacy considerations into software development processes and in integrating the development of safe and secure software. She has published over 100 scientific papers in various fields of software engineering. She is a member of the board of paluno – The Ruhr Institute for Software Technology.

She is a member of the European Workshop on Industrial Computer Systems Reliability, Safety and Security (EWICS). EWICS runs the annual International Conference on Computer Safety, Reliability and Security (SAFECOMP), for which Maritta Heisel is a regular program committee member. She also served as a member of the steering committee of the working group on secure ICT research and innovation established by the European Union in the context of the NIS (Network and Information Security) Public–Private Platform, and is coeditor of a deliverable “Cybersecurity Education Snapshot for workforce development in the EU”. Furthermore, she has been appointed as an independent expert and project proposal evaluator by the European Commission. She was a member of the Network of Excellence on Engineering Future Internet Software Services and Systems (NESSoS), which was funded by the European Union. Furthermore, she was involved in different projects, investigating the derivation of design alternatives based on quality requirements, and developing a support tool for documenting security requirements for cloud computing systems, checking their consistency, and generating documentation for the certification of cloud computing systems. Her personal website is: <https://www.uni-due.de/swe/maritta>.

**Bruce Maxim** is Associate Professor of Computer and Information Science at the University of Michigan-Dearborn. He has worked as a software engineer, project manager, educator, author, and consultant for more than 30 years. His professional experience includes managing research information systems at the University of Michigan Medical School, directing instructional computing for the University medical campus, and working as a statistical programmer in the School of Public Health. He served as the chief technology officer for a game development company. His research interests include software engineering, human computer interaction, game design, social media, artificial intelligence, and software engineering education. He is coauthor of nine books, including a best-selling introductory computer science text and a best-selling software engineering text. He has published a number of papers on software quality, project management, computer algorithm animation, game development, and engineering education. He is the chief-architect of the ABET accredited software engineering program at University of Michigan-Dearborn. He established the GAME Lab in the University of Michigan-Dearborn College of Engineering and Computer Science. He has supervised several hundred industry-based software development projects at both the graduate and undergraduate levels. He is the recipient of several distinguished teaching awards and a distinguished community service award. He is a member of Sigma Xi, Upsilon Pi Epsilon, Pi Mu Epsilon, Association of Computing Machinery, IEEE Computer Society, American Society for Engineering Education, Society of Women Engineers, and International Game Developers Association. His personal website is: <http://www-personal.umd.umich.edu/~bmaxim/>.

