Edward Curry · Sören Auer ·
Arne J. Berre · Andreas Metzger ·
Maria S. Perez · Sonja Zillner *Editors*

Technologies and Applications for Big Data Value

BIG DATA VALUE
PUBLIC-PRIVATE PARTNERSHIP

OPEN ACCESS

Technologies and Applications for Big Data Value

Edward Curry • Sören Auer • Arne J. Berre •
Andreas Metzger • Maria S. Perez • Sonja Zillner
Editors

Technologies and Applications for Big Data Value

Editors
Edward Curry
Insight SFI Research Centre for Data Analytics
NUI Galway, Ireland

Sören Auer
Information Centre for Science and Technology
Leibniz University Hannover
Hannover, Germany

Arne J. Berre
SINTEF Digital
Oslo, Norway

Andreas Metzger
Paluno
University of Duisburg-Essen
Essen, Germany

Maria S. Perez
Universidad Politécnica de Madrid
Boadilla del Monte
Madrid, Spain

Sonja Zillner
Siemens Corporate Technology
München, Germany

ISBN 978-3-030-78306-8 ISBN 978-3-030-78307-5 (eBook)
https://doi.org/10.1007/978-3-030-78307-5

© The Editor(s) (if applicable) and The Author(s) 2022. This book is an open access publication.
Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.
The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Computer science was created by humankind to solve problems. In 100 BC, early hand-powered computing devices such as the Antikythera mechanism were designed to calculate astronomical positions. In the 1800s, Charles Babbage proposed the Analytical Engine to solve general-purpose computational tasks. In the 1900s, the Bomde by Turing and Welchman was critical to code-breaking. Advances in computer science have been driven by the need for humanity to solve the most pressing challenges of the day. Today, computer science tackles significant societal challenges like organising the world's information, personalised medicine, the search of the Higgs boson, climate change, and weather forecasts.

This book aims to educate the reader on how recent advances in technologies, methods, and processes for big data and data-driven Artificial Intelligence (AI) can deliver value to address problems in real-world applications. The book explores cutting-edge solutions and best practices for big data and data-driven AI and applications for the data-driven economy. It provides the reader with a basis for understanding how technical issues can be overcome to offer real-world solutions to major industrial areas, including health, energy, transport, finance, manufacturing, and public administration.

The book's contributions emanate from the Big Data Value Public-Private Partnership (BDV PPP) and the Big Data Value Association, which have acted as the European data community's nucleus to bring together businesses with leading researchers to harness the value of data to benefit society, business, science, and industry. The technological basis established in the BDV PPP will seamlessly enable the European Partnership on AI, Data, and Robotics.

The book is of interest to two primary audiences: first, undergraduate and postgraduate students and researchers in various fields, including big data, data science, data engineering, machine learning, and AI. Second, practitioners and industry experts engaged in data-driven systems and software design and deployment projects who are interested in employing these advanced methods to address real-world problems.

This book is arranged in two parts. The first part contains "horizontal" contributions of technologies and methods which can be applied in any sector. The second

part includes contributions of innovative processes and applications within specific "vertical" sectors. Chapter 1 provides an overview of the book by positioning the chapters in terms of their contributions to technology frameworks, including the Big Data Value Reference Model and the European AI, Data and Robotics Framework, which are key elements of the BDV PPP and the Partnership on AI, Data and Robotics.

Part I: Technologies and Methods details key technical contributions which enable data value chains. Chapter 2 investigates ways to support semantic data enrichment at scale. The trade-offs and challenges of serverless data analytics are examined in Chap. 3. Benchmarking of big data and AI pipelines is the objective of Chap. 4, while Chap. 5 presents an elastic software architecture for extreme-scale big data analytics. Chapter 6 details privacy-preserving technologies for trusted data spaces. Leveraging data-driven infrastructure management to facilitate AIOps is the focus of Chap. 7, and unified big-data workflows over High-Performance-Computing (HPC) and the cloud are tackled in Chap. 8.

Part II: Processes and Applications details experience reports and lessons from using big data and data-driven approaches in processes and applications. The chapters are co-authored with industry experts and cover domains including health, law, finance, retail, manufacturing, mobility, and smart cities. Chapter 9 presents a toolkit for deep learning and computer vision over HPC and cloud architectures. Applying AI to manage acute and chronic clinical conditions is the focus of Chap. 10, while Chap. 11 explores 3D human big data exchange between the health and garment sectors. In Chap. 12, we see how legal knowledge graphs can be used for multilingual compliance services in labour law, contract management, and geothermal energy. Chapter 13 focuses on big data analytics in the banking sector with guidelines and lessons learned from CaixaBank. Chapter 14 explores data-driven AI and predictive analytics for the maintenance of industrial machinery using digital twins. Chapter 15 investigates big data analytics in the manufacturing sector, and Chap. 16 looks at the next generation of data-driven factory operations and optimisation. Large-scale trials of data-driven service engineering are covered in Chap. 17. Chapter 18 describes approaches for model-based engineering and semantic interoperability for digital twins across the product life cycle. In Chap. 19, a data science pipeline for big linked earth observation data is presented, and Chap. 20 looks ahead towards cognitive ports of the future. Distributed big data analytics in a smart city is the focus of Chaps. 21, and 22 looks at system architectures and applications of big data in the maritime domain. The book closes with Chap. 23 exploring knowledge modelling and incident analysis for cargo.

Galway, Ireland
January 2022

Edward Curry

Acknowledgements

The editors thank Ralf Gerstner and all at Springer for their professionalism and assistance throughout the journey of this book. This book was made possible through funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 732630 (BDVe). This work was supported by the Science Foundation Ireland, co-funded by the European Regional Development Fund under Grant SFI/12/RC/2289_P2.

We thank all the authors for sharing their work through our book. Very special thanks to all the reviewers who gave their time, effort, and constructive comments that enhanced the overall quality of the chapters. We particularly recognise the dedication and commitments of Amin Anjomshoaa, Felipe Arruda Pontes, Alexandru Costan, Praneet Dhingra, Jaleed Khan, Manolis Marazakis, Ovidiu-Cristian Marcu, Niki Pavlopoulou, Emilio Serrano, Atiya Usmani, Piyush Yadav, and Tarek Zaarour as reviewers.

Finally, we would like to thank our partners at the European Commission, in particular Commissioner Gabriel, Commissioner Kroes, and the Director-General of DG CONNECT Roberto Viola, who had the vision and conviction to develop the European data economy. We thank the current and past members of the European Commission's Unit for Data Policy and Innovation (Unit G.1), Yvo Volman, Márta Nagy-Rothengass, Kimmo Rossi, Beatrice Covassi, Stefano Bertolo, Francesco Barbato, Wolfgang Treinen, Federico Milani, Daniele Rizzi, and Malte Beyer-Katzenberger. Together they have represented the public side of the Big Data Value Partnership and were instrumental in its success.

January 2022

Edward Curry
Sören Auer
Arne J. Berre
Andreas Metzger
Maria S. Perez
Sonja Zillner

Contents

Technologies and Applications for Big Data Value 1
Edward Curry, Sören Auer, Arne J. Berre, Andreas Metzger, Maria S. Perez, and Sonja Zillner

Part I Technologies and Methods

Supporting Semantic Data Enrichment at Scale 19
Michele Ciavotta, Vincenzo Cutrona, Flavio De Paoli, Nikolay Nikolov, Matteo Palmonari, and Dumitru Roman

Trade-Offs and Challenges of Serverless Data Analytics 41
Pedro García-López, Marc Sánchez-Artigas, Simon Shillaker, Peter Pietzuch, David Breitgand, Gil Vernik, Pierre Sutra, Tristan Tarrant, Ana Juan-Ferrer, and Gerard París

Big Data and AI Pipeline Framework: Technology Analysis from a Benchmarking Perspective 63
Arne J. Berre, Aphrodite Tsalgatidou, Chiara Francalanci, Todor Ivanov, Tomas Pariente-Lobo, Ricardo Ruiz-Saiz, Inna Novalija, and Marko Grobelnik

An Elastic Software Architecture for Extreme-Scale Big Data Analytics 89
Maria A. Serrano, César A. Marín, Anna Queralt, Cristovao Cordeiro, Marco Gonzalez, Luis Miguel Pinho, and Eduardo Quiñones

Privacy-Preserving Technologies for Trusted Data Spaces 111
Susanna Bonura, Davide Dalle Carbonare, Roberto Díaz-Morales, Marcos Fernández-Díaz, Lucrezia Morabito, Luis Muñoz-González, Chiara Napione, Ángel Navia-Vázquez, and Mark Purcell

Leveraging Data-Driven Infrastructure Management to Facilitate AIOps for Big Data Applications and Operations 135
Richard McCreadie, John Soldatos, Jonathan Fuerst, Mauricio Fadel Argerich, George Kousiouris, Jean-Didier Totow, Antonio Castillo Nieto, Bernat Quesada Navidad, Dimosthenis Kyriazis, Craig Macdonald, and Iadh Ounis

Leveraging High-Performance Computing and Cloud Computing with Unified Big-Data Workflows: The LEXIS Project 160
Stephan Hachinger, Martin Golasowski, Jan Martinovič, Mohamad Hayek, Rubén Jesús García-Hernández, Kateřina Slaninová, Marc Levrier, Alberto Scionti, Frédéric Donnat, Giacomo Vitali, Donato Magarielli, Thierry Goubier, Antonio Parodi, Andrea Parodi, Piyush Harsh, Aaron Dees, and Olivier Terzo

Part II Processes and Applications

The DeepHealth Toolkit: A Key European Free and Open-Source Software for Deep Learning and Computer Vision Ready to Exploit Heterogeneous HPC and Cloud Architectures 183
Marco Aldinucci, David Atienza, Federico Bolelli, Mónica Caballero, Iacopo Colonnelli, José Flich, Jon A. Gómez, David González, Costantino Grana, Marco Grangetto, Simone Leo, Pedro López, Dana Oniga, Roberto Paredes, Luca Pireddu, Eduardo Quiñones, Tatiana Silva, Enzo Tartaglione, and Marina Zapater

Applying AI to Manage Acute and Chronic Clinical Condition 203
Rachael Hagan, Charles J. Gillan, and Murali Shyamsundar

3D Human Big Data Exchange Between the Healthcare and Garment Sectors 225
Juan V. Durá Gil, Alfredo Remon, Iván Martínez Rodriguez, Tomas Pariente-Lobo, Sergio Salmeron-Majadas, Antonio Perrone, Calina Ciuhu-Pijlman, Dmitry Znamenskiy, Konstantin Karavaev, Javier Ordono Codina, Laura Boura, Luísa Silva, Josep Redon, Jose Real, and Pietro Cipresso

Using a Legal Knowledge Graph for Multilingual Compliance Services in Labor Law, Contract Management, and Geothermal Energy 253
Martin Kaltenboeck, Pascual Boil, Pieter Verhoeven, Christian Sageder, Elena Montiel-Ponsoda, and Pablo Calleja-Ibáñez

Big Data Analytics in the Banking Sector: Guidelines and Lessons Learned from the CaixaBank Case 273
Andreas Alexopoulos, Yolanda Becerra, Omer Boehm, George Bravos, Vasilis Chatzigiannakis, Cesare Cugnasco, Giorgos Demetriou, Iliada Eleftheriou, Lidija Fodor, Spiros Fotis, Sotiris Ioannidis, Dusan Jakovetic, Leonidas Kallipolitis, Vlatka Katusic, Evangelia Kavakli, Despina Kopanaki, Christoforos Leventis, Mario Maawad Marcos, Ramon Martin de Pozuelo, Miquel Martínez, Nemanja Milosevic, Enric Pere Pages Montanera, Gerald Ristow, Hernan Ruiz-Ocampo, Rizos Sakellariou, Raül Sirvent, Srdjan Skrbic, Ilias Spais, Giorgos Vasiliadis, and Michael Vinov

Data-Driven Artificial Intelligence and Predictive Analytics for the Maintenance of Industrial Machinery with Hybrid and Cognitive Digital Twins 299
Perin Unal, Özlem Albayrak, Moez Jomâa, and Arne J. Berre

Big Data Analytics in the Manufacturing Sector: Guidelines and Lessons Learned Through the Centro Ricerche FIAT (CRF) Case ... 321
Andreas Alexopoulos, Yolanda Becerra, Omer Boehm, George Bravos, Vassilis Chatzigiannakis, Cesare Cugnasco, Giorgos Demetriou, Iliada Eleftheriou, Spiros Fotis, Gianmarco Genchi, Sotiris Ioannidis, Dusan Jakovetic, Leonidas Kallipolitis, Vlatka Katusic, Evangelia Kavakli, Despina Kopanaki, Christoforos Leventis, Miquel Martínez, Julien Mascolo, Nemanja Milosevic, Enric Pere Pages Montanera, Gerald Ristow, Hernan Ruiz-Ocampo, Rizos Sakellariou, Raül Sirvent, Srdjan Skrbic, Ilias Spais, Giuseppe Danilo Spennacchio, Dusan Stamenkovic, Giorgos Vasiliadis, and Michael Vinov

Next-Generation Big Data-Driven Factory 4.0 Operations and Optimization: The Boost 4.0 Experience 345
Oscar Lázaro, Jesús Alonso, Philip Ohlsson, Bas Tijsma, Dominika Lekse, Bruno Volckaert, Sarah Kerkhove, Joachim Nielandt, Davide Masera, Gaetano Patrimia, Pietro Pittaro, Giuseppe Mulè, Edoardo Pellegrini, Daniel Köchling, Thanasis Naskos, Ifigeneia Metaxa, Salome Leßmann, and Sebastian von Enzberg

Big Data-Driven Industry 4.0 Service Engineering Large-Scale Trials: The Boost 4.0 Experience 373
Oscar Lázaro, Jesús Alonso, Paulo Figueiras, Ruben Costa, Diogo Graça, Gisela Garcia, Alessandro Canepa, Caterina Calefato, Marco Vallini, Fabiana Fournier, Nathan Hazout, Inna Skarbovsky, Athanasios Poulakidas, and Konstantinos Sipsas

Model-Based Engineering and Semantic Interoperability for Trusted Digital Twins Big Data Connection Across the Product Lifecycle 399
Oscar Lázaro, Jesús Alonso, Roxana-Maria Holom, Katharina Rafetseder, Stefanie Kritzinger, Fernando Ubis, Gerald Fritz, Alois Wiesinger, Harald Sehrschön, Jimmy Nguyen, Tomasz Luniewski, Wojciech Zietak, Jerome Clavel, Roberto Perez, Marlene Hildebrand, Dimitris Kiritsis, Hugues-Arthur Garious, Silvia de la Maza, Antonio Ventura-Traveset, Juanjo Hierro, Gernot Boege, and Ulrich Ahle

A Data Science Pipeline for Big Linked Earth Observation Data 431
Manolis Koubarakis, Konstantina Bereta, Dimitris Bilidas, Despina-Athanasia Pantazi, and George Stamoulis

Towards Cognitive Ports of the Futures 453
Santiago Cáceres, Francisco Valverde, Carlos E. Palau, Andreu Belsa Pellicer, Christos A. Gizelis, Dimosthenes Krassas, Hanane Becha, Réda Khouani, Andreas Metzger, Nikos Tzagkarakis, Anthousa Karkoglou, Anastasios Nikolakopoulos, Achilleas Marinakis, Vrettos Moulos, Antonios Litke, Amir Shayan Ahmadian, and Jan Jürjens

Distributed Big Data Analytics in a Smart City 475
Maria A. Serrano, Erez Hadad, Roberto Cavicchioli, Rut Palmero, Luca Chiantore, Danilo Amendola, and Eduardo Quiñones

Processing Big Data in Motion: Core Components and System Architectures with Applications to the Maritime Domain 497
Nikos Giatrakos, Antonios Deligiannakis, Konstantina Bereta, Marios Vodas, Dimitris Zissis, Elias Alevizos, Charilaos Akasiadis, and Alexander Artikis

Knowledge Modeling and Incident Analysis for Special Cargo 519
Vahideh Reshadat, Tess Kolkman, Kalliopi Zervanou, Yingqian Zhang, Alp Akçay, Carlijn Snijder, Ryan McDonnell, Karel Schorer, Casper Wichers, Thomas Koch, Elenna Dugundji, and Eelco de Jong