

Book review: Textbook series on big data management and applications

The Innovation Informatics Editorial Team, China*

*Correspondence: informatics@the-innovation.org

Received: January 20, 2026; Accepted: February 10, 2026; Published Online: February 14, 2026; <https://doi.org/10.59717/j.xinn-inform.2026.100035>

© 2026 The Author(s). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

Citation: The Innovation Informatics Editorial Team, China (2026). Book review: Textbook series on big data management and applications. *The Innovation Informatics* 2:100035.

Big data analytics is undergoing a profound paradigm shift—from passive data storage to agentic intelligence—fundamentally reshaping knowledge transmission systems in informatics. Despite this transformation, the growing demand for comprehensive and integrated training in big data mining and intelligent applications among students and practitioners has not yet been fully met. To address this gap, a research team led by Professor Xin Lu of the National University of Defense Technology initiated the development of a comprehensive textbook series on Big Data Management and Applications.

THE EVOLVING LANDSCAPE OF BIG DATA ANALYTICS TECHNOLOGY

The global big data ecosystem is transitioning from scale-oriented expansion toward intelligent governance, making “digitalization and intelligence” a defining trend of the digital era. At the data level, previously fragmented and rudimentary data accumulation practices are being standardized and refined, positioning data governance as a central topic in the digital economy. At the algorithmic level, advanced intelligent approaches—such as generative pretrained large models—are playing an increasingly prominent role in fully mining the latent value of big data. Correspondingly, big data analytics technologies have evolved rapidly, enabling more agile data management, deeper insight into complex data patterns, more reliable decision support, and faster

workflow optimization.

In response to this technological revolution and industrial transformation, academic programs in data science and big data technology have expanded rapidly since the concept gained prominence around 2016. By the end of 2025, more than 800 universities across China offered degree programs in data science and big data technology, big data management and applications, and data computation and data-driven engineering.

STRATEGIC BLUEPRINT OF THE TEXTBOOK SERIES

Despite widespread institutional adoption, big data remains a challenging discipline to master. Students frequently encounter large-scale, real-world datasets without systematic training in data analysis, making it difficult to progress beyond conceptual understanding. Although the transformative potential of big data is widely recognized, many learners struggle to extract meaningful patterns, actionable insights, or decision-relevant intelligence from complex data environments. This textbook series seeks to address these challenges by establishing a scalable and structured standard for data science education, organized into a three-tiered architecture: fundamental techniques, core methodologies, and analytical skills (Figure 1).

Big Data Architecture, which constitutes the foundational tier, enables



Figure 1. Strategic framework for the design of Lu's textbook series.

readers to rapidly acquire the capacity to deploy and operate big data platforms through extensive practical cases and supporting resources. It provides in-depth coverage of distributed systems theory, mainstream processing frameworks such as Hadoop, and core technologies including HDFS, Hive, HBase, MapReduce, Spark, and Flink, as well as system tools such as ZooKeeper, YARN, Flume, Kafka, TiDB, Loki, and Milvus.

At the level of core methodologies, *Data Mining* systematically presents essential modeling and analytical techniques, including data description and statistical indicators, linear and nonlinear correlation analysis, regression analysis, dimensionality reduction, association rule mining, classification, clustering, anomaly detection, and ensemble learning. Building on this foundation, *Machine Learning and Artificial Intelligence* guides readers in constructing a comprehensive knowledge system of modern AI, covering core principles of machine learning, intelligent optimization, neural networks and deep learning, reinforcement learning, transfer learning, federated learning, and related cutting-edge applications.

Two major branches of big data analytics—*Natural Language Processing* (NLP) and *Complex Networks*—are also integral components of the series. The NLP volume introduces fundamental concepts, text representation methods, pretraining techniques, and large language models, alongside practical applications such as Chinese word segmentation, part-of-speech tagging, named entity recognition, syntactic parsing, topic modeling, text clustering, sentiment analysis, and knowledge graph construction. The *Complex Networks* volume focuses on relational data mining and system-level analysis, covering graph theory and network science, network metrics and models, sampling methods, robustness analysis, link prediction, informa-

tion propagation, community detection, temporal and higher-order networks, and graph representation learning.

The final tier, analytical skills, includes *Big Data Practice with Python* and *Big Data Visualization*. The former uses Python as a unifying tool to integrate programming fundamentals, big data platform operations, database management, web crawling, data mining, and deep learning, thereby strengthening readers' practical modeling and analytical capabilities. The latter introduces the principles and science of data visualization, enabling readers to transform complex data into intuitive and effective visual representations. Topics include visualization fundamentals, general and specialized tools, design principles, and visualization techniques for structured data, text data, network data, geographic data, and other complex data types.

IMPACT AND OUTLOOK

After five years of sustained effort, *Data Mining* and *Big Data Architecture*, two flagship volumes of the series, have been published and adopted as core undergraduate and postgraduate textbooks at more than 30 universities. The remaining volumes—*Machine Learning and Artificial Intelligence*, *Natural Language Processing*, *Complex Networks*, *Big Data Practice with Python*, and *Big Data Visualization*—are expected to be released later this year, further consolidating the series as a comprehensive educational framework for the next generation of data scientists.

DECLARATION OF INTERESTS

The author declares no competing interests.