Matloff, Norman
Parallel computing for data science. With examples in R, C++ and CUDA. (English)
Zbl 1330.68010
Chapman & Hall/CRC The R Series. Boca Raton, FL: CRC Press (ISBN 978-1-4665-8701-4/hbk). xxiii, 324 p. (2016).

The book is a modern handbook of parallel computing for data science. For this reason, it provides a dozen of algorithms in the R, C++ and CUDA languages for solving many problems in the data sciences, including statistics, data mining and machine learning, pattern recognition, etc. The book consists of thirteen chapters. Chapter 1 presents an introductory code example in parallel processing. Chapter 2 explains what the general factors are that can reduce a parallel program's speed. Principles of parallel loop scheduling are presented in Chapter 3. In Chapters 4–7 the shared-memory paradigms of parallel processing in the programming languages are discussed. The message-passing paradigm is given in Chapter 8. MapReduce computation, parallel sorting and merging are widely explained in the two successive chapters. Examples of parallel filter operation, parallel computation of cumulative sums etc., using all the languages mentioned above, are demonstrated in Chapter 11. Parallel matrix operations for solving various problems, such as graph connectedness, Gaussian elimination and the LU decomposition, the Jacobi algorithm, QR decomposition etc., are outlined in Chapter 12. Methods for converting many non-embarrassingly parallel (EP) applications into statistically equivalent EP substitutes for reducing their computational complexity are treated in Chapter 13.

In summary, this book provides code for many examples in the data sciences, including statistics, data mining and machine learning. It explains how many software packages ease the programming of multicore machines and the graphics processing unit (GPU). All examples are poignant, and the presentation makes the contents easily accessible to all readers, including researchers and students in a wide range of disciplines.

Reviewer: Jerzy Martyna (Kraków)

MSC:

68-02	Research exposition (monographs, survey articles) pertaining to computer science
62H30	Classification and discrimination; cluster analysis (statistical aspects)
68N15	Theory of programming languages
68N19	Other programming paradigms (object-oriented, sequential, concurrent, automatic, etc.)
68P05	Data structures
68T05	Learning and adaptive systems in artificial intelligence
68T10	Pattern recognition, speech recognition
68W10	Parallel algorithms in computer science

Cited in **1** Document

Keywords:

software; data structures; programming languages

Software:

Rdsm; matpow; R; bigmemory; dbgR; partools; CUDA; snow; CRAN; Rmpi