

5 FANTASTIC BOOKS TO READ ON BIG DATA AND DATA SCIENCE



If you've been following this [6-part series on big data](#), then you have learned quite a bit about what big data is, how it can benefit your business, and the things to look for when hiring big data talent. For the last and final post in this series, I want to leave you with some recommendations for additional materials that will help continue to guide you on your big data journey.

Big Data: A Revolution that will Transform how we Live, Work and Think

By: Viktor Mayer-Schonberger and Kenneth Cukier

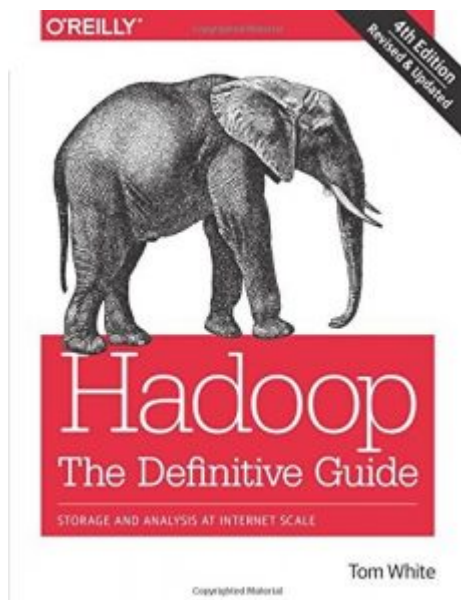


[*Big Data: A Revolution*](#) provides a broad overview of big data and the impact that it's making on modern society. While it's by no means a technical book, it does provide a good high-level introduction to what big data is and how it's affecting practices in areas as diverse as fraud detection and international law enforcement, to linguistics and automated language translation. Well-suited for business managers and analysts, or maybe even C-level executives, *Big Data: A Revolution* provides insight and guidance on how industries should move forward in the wake of today's information revolution.

One of the book's central premises is the notion of "why, not *what*". For example, the book states, "the era of big data challenges the way we live and interact with the world. Most strikingly, society will need to shed some of its obsession for causality in exchange for simple correlations: not knowing why but only what." Later in the book, authors reemphasize this point when they write, "Big data is about what, not why. We don't always need to know the cause of a phenomenon, rather, we can let the data speak for itself," emphasizing that "knowing what, not why is good enough." *Excerpts from - [Big Data: A Revolution that will Transform how we Live, Work and Think](#)*

Hadoop: The Definitive Guide, 4th Edition

By: Tom White



[*Hadoop: The Definitive Guide*](#) is a big data book that's targeted at

technical audiences. The book was originally published in 2009, and is currently sold as a 4th edition update. Praised by developers and data engineers the world-over, *Hadoop: The Definitive Guide* provides how-to's on building and maintaining distributed, parallel processing data systems with Apache Hadoop (HDFS, MapReduce, and YARN). The 4th Edition update even goes into details on Hadoop 2 deployment, including technical details that you should know about YARN, HBase, Parquet, Flume, Crunch, Pig, Hive, and Spark. The book also presents interesting case studies from the healthcare industries and from genomic sciences.

Data Smart: Using Data Science to Transform Information into Insight

By: John Foreman



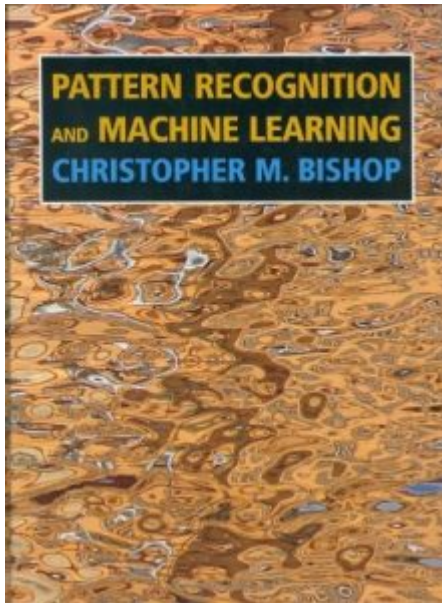
Business professionals love [*Data Smart*](#), it's as simple as that! Written especially for data science newbies, *Data Smart* provides a really easy way for readers to grasp the concepts and techniques that underlie data science. Furthermore, the book provides step-by-step tutorials on how to execute these techniques in simple Excel software. Some data science methods covered in *Data Smart* include:

- Cluster analysis (including k-means and k-medians methods)
- Linear programming for document classification
- Various forms of linear regression analysis
- Time series forecasting

Although this book won't teach you everything you need to know in order to start deploying large-scale analytics projects, it will help you learn the basic ABCs of data science and some of the methods comprising it.

Pattern Recognition and Machine Learning

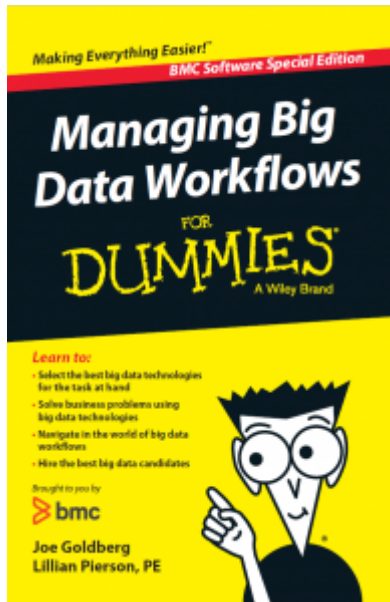
By: Christopher Bishop



Pattern Recognition and Machine Learning is another great book about data science. In contrast to *Data Smart*, however, *Pattern Recognition* was written to satisfy the interests and technical capacities of already advanced information scientists and statisticians. The book introduces inferential approximation algorithms that are useful in generating fast answers from questions asked of big data sets. Although the book requires no prerequisite knowledge of pattern recognition or machine learning, it does specify that readers should be skillful in calculus and the basics of probability and linear algebra. Engineers and statisticians have praised the book for its readability and comprehensiveness, although critics have voiced frustration with its non-intuitive math-heavy approach.

Managing Big Data Workflows for Dummies

By: Joe Goldberg and Lillian Pierson, P.E.



Managing Big Data Workflows for Dummies, recently published, was written for IT decision-makers to provide them an overview of big data, including key technologies, and some common industry applications. One key focus of this book is workload management – the process of automating big data workflows across the enterprise. The book is currently available for free, as an ebook download. [Download here](#).