

**The George Washington University
School of Engineering and Applied Science
Department of Computer Science
CSCI 6444 – Big Data Analytics –**

- Big Data: Principles and Paradigms, by Rajkumar Buyya, Rodrigo N. Calheiros, Amir Vahid Dastjerdi
- Mining of Massive Datasets: <http://www.mmds.org/>
- Data Wrangling: [https://www.bankingtech.com/files/2017/10/Trifacta Principles-of-Data-Wrangling.pdf](https://www.bankingtech.com/files/2017/10/Trifacta_Principles-of-Data-Wrangling.pdf)
- Database System Concepts: <https://www.db-book.com/>
- Use Case Walkthroughs for Real-Time Stream Processing: https://info.decodable.co/ebook-use-case-walkthroughs-for-real-time-stream-processing?utm_source=google&utm_medium=cpc&utm_campaign=Search - Data Stream - General&obility_id=18600955589-147502378047&utm_content=640930299337&utm_term=data%20streaming&gclid=CjwKCAiAqaWdBhAvEiwAGAQLtomr6ntlqoCQ4vWTyMafKTWMoyMt16KCTwRaeFCqSVltZuEJfJq8fxoCpwMQAvD_BwE
- Data-Intensive Text Processing with MapReduce, Jimmy Lin and Chris Dyer (www.umiacs.umd.edu/~jimmylin/MapReduce-book-final.pdf)
- Big Data Principles and Paradigms: http://dphoto.lecturer.pens.ac.id/lecture_notes/internet_of_things/Big%20Data%20Principles%20and%20Paradigms.pdf
- Learning Spark: Lightning-Fast Big Data AnalysisPage1of6
 - Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia
 - ISBN: 978-1449358624
 - Copyright: 2015
 - Publisher: O'Reilly Media
- Hadoop Application Architectures: Designing Real-World Big Data Applications
 - Mark Grover, Ted Malaska, Jonathan Seidman, and Gwen Shapira
 - ISBN: 978-1491900086
 - Copyright: 2015
 - Publisher: O'Reilly Media
- HBase: The Definitive Guide: Random Access to Your Planet-Size Data
 - Lars George
 - ISBN: 978-1449396107
 - Copyright: 2011
 - Publisher: O'Reilly Media
- Principles of Data Wrangling
 - Joseph M. Hellerstein, Tye Rattenbury, Jeffrey Heer, Sean Kandel, Connor Carreras
 - ISBN: 9781491938928
 - Released July 2017
 - Publisher(s): O'Reilly Media, Inc.
- Data-Intensive Text Processing with MapReduce, Jimmy Lin and Chris Dyer: <https://lintool.github.io/MapReduceAlgorithms/MapReduce-book-final.pdf>

- Big Data: Principles and Paradigms, by Rajkumar Buyya, Rodrigo N. Calheiros, Amir Vahid Dastjerdi
- “Streams and Tables: Two Sides of the Same Coin,” Matthias J. Sax, et al, BIRTE '18: Proceedings of the International Workshop on Real-Time Business Intelligence and Analytics, August 2018
- Base Properties: : <http://queue.acm.org/detail.cfm?id=1394128>
- <https://www.mongodb.com/>
- Clustering using MapReduce:
Apache Mahout Essentials, By Jayani Withanawasam
https://web2.qatar.cmu.edu/~mhhammou/15440-f19/recitations/Project4_Handout.pdf
- PLANET: MapReduce Classification
<https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/36296.pdf>
<https://download.atlantis-press.com/article/4283.pdf>