[1] Krishna Sankar and Holden Karau, “Building and Running a Spark Application,” in Fast Data Processing with Spark - Second Edition, 2nd ed. Mumbai, India: Packt Publishing, 2016.

This chapter gives detailed steps for building your Spark project with sbt, Maven or anything else.

[2] Amin Hosseinian-Far, Muthu Ramachandran, Dilshad Sarwar, Strategic Engineering for Cloud Computing and Big Data Analytics, 1st ed. Leeds, UK: Springer International Publishing AG, 2017.

This book describes the cloud services, big data analytics and business process modelling in detail. Also describes the sustainability, security and safety of the applications using cloud.

[3] Shyam R., Bharathi ; Ganesh H.B., Sachin ; Kumar S., Prabaharan ; Poornachandran, Prabaharan ; Soman K.P., Prabaharan, “[Apache Spark a Big Data Analytics Platform for Smart Grid](https://fresnostate-primo.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=TN_sciversesciencedirect_elsevierS2212-0173(15)00313-8&context=PC&vid=01CALS_UFR&lang=en_US&search_scope=articles&adaptor=primo_central_multiple_fe&tab=articles&query=any,contains,apache%20spark,AND&sortby=rank&mode=advanced&offset=0),” in Procedia Technology, Fresno, CA, 2015.

This article describes Apache Spark as an integrated platform for development of smart applications.

[4] Mavridis, Ilias ; Karatza, Helen, “Performance evaluation of cloud-based log file analysis with Apache Hadoop and Apache Spark,” in The Journal of Systems & Software, Berkeley, CA, 2009.

This article describes the *distributed SQL-type querying with****Apache****Hive and****Spark****SQL.*

*[5]* Arias, Jacinto ; Gamez, Jose A. ; Puerta, Jose M., “Learning distributed discrete Bayesian Network Classifiers under Map reduce with Apache Spark,” in Knowledge-Based Systems, Yakima, WA, 2017.

This article describes the state-of –the-art Apache computing framework.

[6] Huang, Wei ; Meng, Lingkui ; Zhang, Dongying ; Zhang, Wen, “In-memory parallel processing of massive remotely sensed data using an Apache Spark on Hadoop YARN Model,” in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Beijing, China, 2017.

[7] Gupta, Govind P. ; Kulariya, Manish, “A framework for fast and efficient cyber security network,” in Procedia Computer Science, 2016.

[8] Zhang, Qiang ; Vatsavai, Ranga ; Shashidharan, “Agent based urban growth on Apache Spark,” in Proceedings of the 5th ACM SIGSPATIAL International Workshop on analytics for big geospatial data, 2016.

[9] Baer, Troy ; Peltz, Paul ; Yin, Junqi ; Begoli, Edmon, “Integrated Apache Spark into HPC environment,” in Proceedings of the 2015 XSEDE Conference, 2015

[10] How Big Data is used in Practice, [online] Available: <https://www.bernardmarr.com/default.asp?contentID=1076>

This online text describes ten major use cases where big data is currently being used and also gives information about some companies that really use big data to improve performance.

[11] Big Data, [online] Available: <https://www.sas.com/en_us/insights/big-data/what-is-big-data.html>

This website gives us an introduction to big data, explains its importance and states its features. It gives a clear understanding of its concept.

# [12]Big Data Processing with Apache Spark, [online] Available: <https://www.infoq.com/articles/apache-spark-introduction>

# This website focusses on the advantages of Apache Spark over the conventional Big Data analytics technique: Hadoop.

[13] How Apache Spark fits into the Big Data landscape, [online] Available: <https://lintool.github.io/SparkTutorial/slides/day1_context.pdf>

This online resource presents information in slides format. It describes the Big Data challenges like collection, ETL, storage, exploration and analytics, and explains why the in-memory performance of spark should be considered.

[14] Apache Spark, [online] Available: <https://spark.apache.org/>

This is the official website of Apache Spark. It contains links to download Apache Spark, its libraries, gives access to its forum and documentation of its releases.

[15] Sandy Ryza, Uri Laserson, Josh Wills and Sean Owen, Advanced Analytics with Spark, 2nd ed. NY: O’Reilly Media, 2017.

This book presents a set of self-contained patterns for performing large-scale data analysis with Spark. The authors bring Spark, statistical methods, and real-world data sets together to explain the working of Spark.