

Useful Resources

Overall instruction on the methods and processes:

- Suthaharan, Shan. Machine learning models and algorithms for big data classification. Boston: Springer, 2016.

Install Ubuntu on a VM in VirtualBox:

- <https://www.youtube.com/watch?v=hK-oggHEetc>

Install Hadoop on Ubuntu (single cluster):

- <https://hadoop.apache.org/docs/r2.7.1/hadoop-project-dist/hadoop-common/SingleCluster.html>
- <https://rsqrl.com/train/train1.html>
- <http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/>

Install R and RStudio Server on Ubuntu:

- <https://cran.r-project.org/bin/linux/ubuntu/README.html>
- <https://www.rstudio.com/products/rstudio/download-server/>

Install RHadoop packages:

- <http://hadooptutorial.info/rhadoop-installation-on-ubuntu/>

Linux Terminal Command Reference:

- <https://community.linuxmint.com/tutorial/view/244>
- <https://help.ubuntu.com/community/UsingTheTerminal>

Hadoop Distributed File Systems and MapReduce Framework and:

- J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," Communications of the ACM, vol. 51, no. 1, pp. 107-113, 2008.
- http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html#Large_Data_Sets

Word Count example

- <https://github.com/jeffreybreen/tutorial-rmr2-airline/blob/master/R/1-wordcount.R>