

Data Science Machine Setup and required Tools

Aritra Biswas

January 12, 2017

How to setup a machine with the open source working tools one need for data science? This is one of the major question among most of the students and enthusiasts who are changing their platform from Windows or Mac to Linux. There is no doubt about the fact that the open source community one of the biggest advantage of R and Python. In order to leverage from other peoples code in a open source platform is not a easy thing to do. Here I'm going to share my experiences that how one can setup a computer with Ubuntu and other important tools for a data science project. Here is a list of tools which are required for this process.

- A 8 GB pen drive or a DVD
- Rufus or Any other boot-able use creator
- A ISO image of Ubuntu
- Good internet connection

These are some of the basic things one needs in order to setup a computer with Ubuntu and after that these are the tools of my choice which can be very useful for data science projects:

- R (Core/Developers)
- R studio
- Latex (Full)
- Python (Anaconda)
- Spark (Version 1.6.2 with Hadoop 2.6)
- JDK
- Notepad++
- Git
- WPS office
- PDF reader
- OCR
- Mingw
- Code Blocks

These are the list of soft wares which are I feel one should have in order to start working on a data science project.

Here I'm going to mention some of the packages , add ins, libraries or extension for these programming languages which are very much important in order to develop something using these open source tools.