Big Data is a growing field and you probably have a lot to learn if you  want to learn about it.I will try to provide the path I took:  
  
**1. Start by Learning a Programming Language:**  
  
If  you want to tackle Big data you should  know Python/Java. If you don't  know both of these start with Python. Just start with the basics- For  loop, Lists, Dictionaries, Iterating through a list and dictionary etc. I  would advice taking this course on edX:[Introduction to Computer Science and Programming Using Python](https://www.edx.org/course/introduction-computer-science-mitx-6-00-1x-6" \t "https://www.quora.com/_blank)  
In the rest of this post I will assume that you went by my suggestion and are using Python.

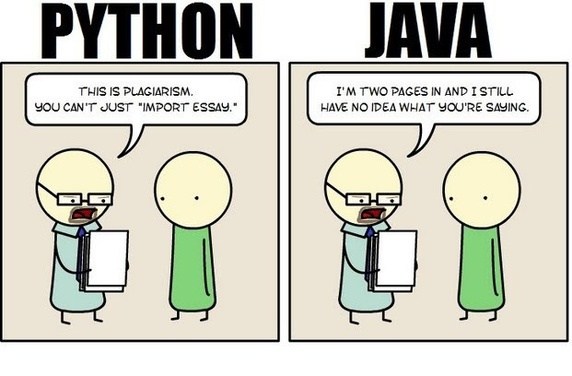


Image Credits: xkcd  
  
**2. Learn about a Big Data Platform:**  
  
Once  you feel that you could solve basic problems using Python/Java, you are  ready for the next step. You need to learn about some Big Data  Technology like Hadoop/Spark. Now you could start with Spark also but I  feel that Hadoop would be the best place to start as it can provide you  with more background of the Mapreduce Paradigm, and you will be able to  understand the problems that introduction of Spark solves.    
To learn Hadoop I would advice you to checkout this course on Udacity:  
[https://www.udacity.com/course/intro-to-hadoop-and-mapreduce--ud617](https://www.udacity.com/course/viewer" \l "!/c-ud617/l-306818608/m-309273649" \t "https://www.quora.com/_blank)  
  
Once  you are done through this course you would have gained quite a basic  understanding of concepts and you would have installed a Hadoop VM in  your own machine. You would also have solved the Basic Wordcount  Problem.   
  
Read this amazing Blog Post from Michael Noll:[Writing An Hadoop MapReduce Program In Python - Michael G. Noll](http://www.michael-noll.com/tutorials/writing-an-hadoop-mapreduce-program-in-python/" \t "https://www.quora.com/_blank).  Just read the basic mapreduce codes. Don't use Iterators and Generators  yet. This has been a starting point for many of us Hadoop developers.  
  
Now try to **solve these two problems** from the CS109 Harvard course from 2013:

A. First, grab the file word\_list.txt from [Page on github.com](https://raw.github.com/cs109/content/master/labs/lab8/word_list.txt" \t "https://www.quora.com/_blank).  This contains a list of six-letter words. To keep things simple, all of  the words consist of lower-case letters only.Write a mapreduce job that  finds all anagrams in word\_list.txt.  
  
B. For the next problem, download the file [baseball\_friends.csv](https://raw.github.com/cs109/content/master/labs/lab8/baseball_friends.csv" \t "https://www.quora.com/_blank). Each row of this csv file contains the following:

* A person's name
* The team that person is rooting for -- either "Cardinals" or "Red Sox"
* A list of that person's friends, which could have arbitrary length

For  example: The first line tells us that Aaden is a Red Sox friend and he  has 65  friends, who are all listed here. For this problem, it's safe to  assume  that all of the names are unique and that the friendship  structure is  symmetric (*i.e.* if Alannah shows up in Aaden's friends list, then Aaden will show up in Alannah's friends list).  
Write  an mr job that lists each person's name, their favorite  team, the  number of Red Sox fans they are friends with, and the number  of  Cardinals fans they are friends with.

Try  to do this yourself. Don't use the mrjob (pronounced Mr. Job) way that  they use in the CS109 class. Use the proper Hadoop Streaming way as  taught in the Udacity class as it is much more customizable in the long  run. If you face problems I could guide you. Ping me up.  
  
If you are done with these, you can safely call yourself as someone who could **"think in Mapreduce"** as how people like to call it.Try to do groupby, filter and joins using Hadoop. You can read up some good tricks from my blog:  
[Hadoop Mapreduce Streaming Tricks and Techniques](http://mlwhiz.com/blog/2015/05/09/Hadoop_Mapreduce_Streaming_Tricks_and_Techniques/" \t "https://www.quora.com/_blank)  
  
And don't forget about the [Hadoop Streaming API. Read it!!!!!](https://hadoop.apache.org/docs/r1.2.1/streaming.html" \t "https://www.quora.com/_blank)  
  
**3. Learn a Little Bit of Bash Scripting:**  
  
In the meantime while you are learning Hadoop and in the process of getting your hands dirty with coding, try to read up on shell scripting.   
It allows you to do simple data related tasks in the terminal itself.  
Read these tutorials for doing that:  
  
[Shell Basics every Data Scientist Should know -Part I](http://mlwhiz.com/blog/2015/10/09/shell_basics_for_data_science/" \t "https://www.quora.com/_blank)  
[Shell Basics every Data Scientist Should know - Part II(AWK)](http://mlwhiz.com/blog/2015/10/11/shell_basics_for_data_science_2/" \t "https://www.quora.com/_blank)  
  
I use shell commands because they are fast and I don't need to write a script for each and everything.  
  
**4. Learn Spark:**



Now  comes the next part of your learning process. This should be undertaken  after a   
little bit of experience with Hadoop. Spark Will provide you  with the speed and tools that Hadoop couldn't. But you need to know  Scala/Python to use it. That is one of the reason I suggested that you  go with Python if you don't know any of Java/Python.   
  
Now  Spark is used for data preparation as well as Machine learning  purposes. I would encourage you to take a look at these two courses on  edX provided by Berkeley instructors. The second course would get you  started a little bit in Machine learning too.  
  
1. [Introduction to Big Data with Apache Spark](https://www.edx.org/course/introduction-big-data-apache-spark-uc-berkeleyx-cs100-1x" \t "https://www.quora.com/_blank)  
2. [Scalable Machine Learning](https://www.edx.org/course/scalable-machine-learning-uc-berkeleyx-cs190-1x" \t "https://www.quora.com/_blank)  
  
I have written a little bit about Basic data processing with spark here:   
[Learning Spark using Python: Basics and Applications](http://mlwhiz.com/blog/2015/09/07/Spark_Basics_Explained/" \t "https://www.quora.com/_blank)  
  
If  you don't go through the courses, try solving the same two problems  above that you solved by Hadoop using Spark too. Otherwise the problem  sets in the two courses are more than enough.  
  
And sorry for all the shameless plugs, but I do feel they add value, so I added them.  
  
Hope this Helps. Now get working!!!