Assignment 1

Bhavneet Soni

Solution 1: In today's world when you think about big data application there is no dearth of every day applications that are driven by Big Data. Daily, we come across numerous big data applications from banking, education, healthcare, transportation, communications, shopping to entertainment that are. Industries are making use of big data to understand and predict the needs of their customers.
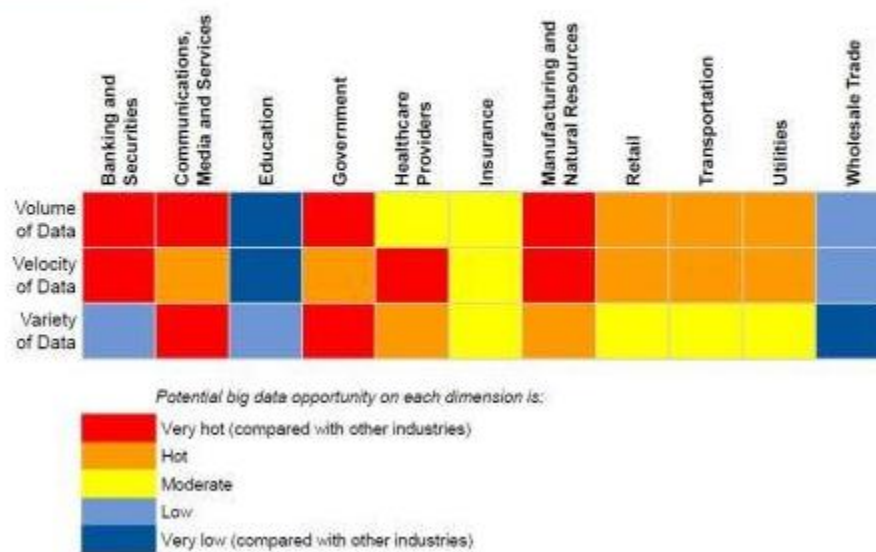

Figure 1: Big Data usage by industry [1]

For this assignment, I want to discuss about how big data is used by Amazon. Anyone who has ever shopped online would have either used Amazon either to buy or just to review stuff. Amazon started off just as an online bookstore but had grew to become an application which everyone goes to compare and review stuff before buying. It is estimated Amazon has over 152 million customers and it draws on an average 1 MB of data from each of the object hosted on its platform. Amazon has access to about 1 exabyte of data i.e. 1 billion gigabytes. The data varies from customer's profile, demographics, education level, search history, preferences to product information, images etc.

As of Jan, 04 2017 Amazon.com has a total of 398,040,250 products for sale, from all sorts of product categories, below is a graph showing top 10 categories by products available at site.
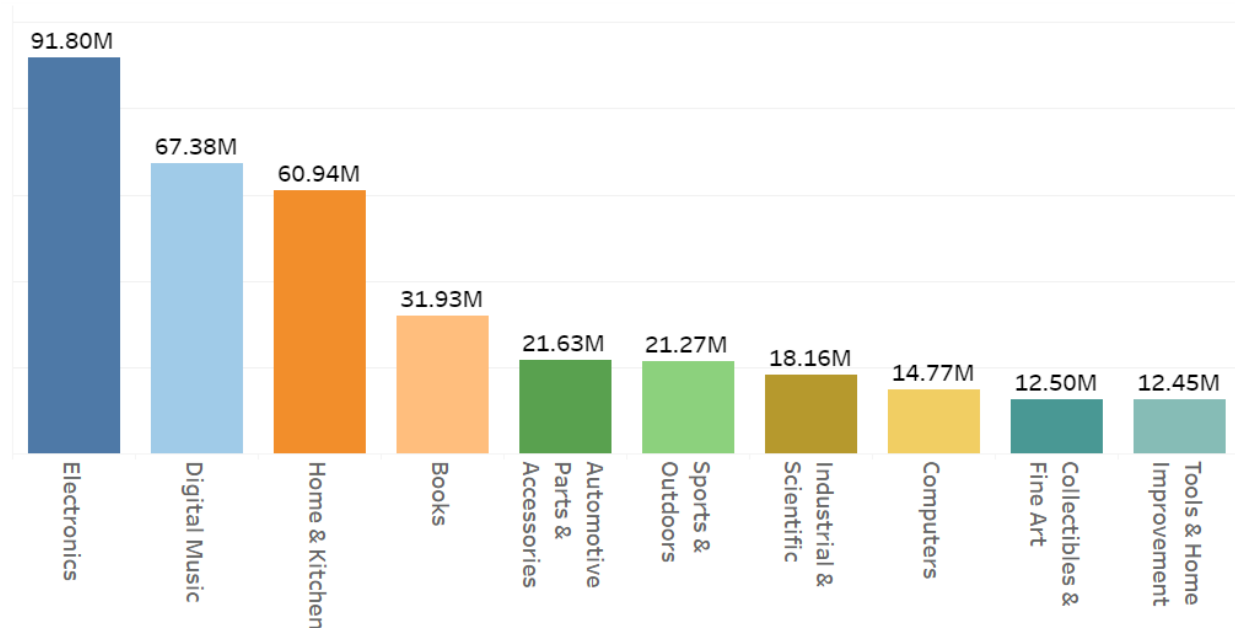

Figure 2: Top 10 Product Categories [2]

With big data comes big issues, like storing data, managing it, backing it up, keeping the data safe and above all extracting meaningful information from it. With high amount of data, the traditional RDBMs are no longer effective and take forever to analyze. Also since the data is of so many different types, traditional databases can't effectively handle them. To manage this huge data, we have effective solutions that are widely used in the industry. To effectively store and to make sure we have sufficient back up of the data we use Distributed File system (DFS), in which the data is broken up in to small packages and multiple copies of these packages are stored in multiple machines, to ensure data backup and data availability in case of a machine failure. Apart from that DFS provides platform for running analytics functions in parallel using MapReduce Algorithm there by giving a fast output. This ability to access and analyses large amounts of data has opened

a lot of venues for the company, they are better able to predict and provide suggestions to their customers, which increase the targeting of the advertisement so much effective. Amazon has gone so big in this that they have started offering big data analytics services to other businesses and applications. However, there are still a few challenges that they face and must overcome in the future to truly utilize the potential of big data.

Solution 2: A 57-bit key will have a combination of 2 to the power of 57 which is about 144,115,188,075,855,872 combinations. A brute force approach is in which all the possible combinations are tried one by one. Using this approach, the worst-case scenario would be when we get the correct on the last try.

So, for a PC capable of checking 1000 combination per millisecond will be able to try about 1000 x 1000 combinations per second, i.e. it will take 144,115,188,076 seconds which is about 40,031,997 hours (144,115,188,076 /60*60). So, it will take about 4570 years to decrypt the key.

While a super computer that can do 10^6 decryptions per microsecond i.e. $10^6 * 10^6$ combinations per sec, so it will take 144,116 seconds ($2^{57} / (10^6 * 10^6)$). which comes out to be a little over 40 hours.

References

[1] M. Gaitho, "How Applications of Big Data Drive Industries," 20 October 2015. [Online]. Available: https://www.simplilearn.com/big-data-applications-in-industries-article.

[2] SCRAPEHERO, "Number of Products Sold on Amazon.com," 5 January 2017. [Online]. Available: https://www.scrapehero.com/how-many-products-are-sold-on-amazon-com-january-2017-report/.