An Era of Big Data
All about Big Data and its Impact on the Current Business

[1]Rashmi S, [2]Hanumanthappa M
[1]Assistant Professor, [2]Professor
[1]Department of Computer Science,
[1]Kristu Jayanti College, Bengaluru – 560077, Karnataka, India

Abstract : Big data, interestingly not a new term, yet groundbreaking terminology in the current world, has transformed many aspects in the business scenarios. There is a tremendous amount of data online, most of which are actually dumped in archives and not being used for the right purpose. In this course of action, big data plays a vital role. The data from the users comes in many ways; structured and unstructured, numbers, images, videos and so on. A recent research study shows that 85% of the data is actually unstructured; I mean look at the amount of images, videos and gif what we sent in what's app or share it on any of the social networking site. It's abundant!!!! Data shared online is increasing every second at an alarming rate. So big data prominently handles these data and analyse it in such a way that it helps organizations to make better business decisions. In this article, a brief analysis on big data is portrayed and a small survey is conducted to gauge the amount of data that is getting accumulated on a day basis. The challenging questions are; what happens to all these data that gets bundled and piled up on the internet, what is the actual use of these data and is it being used for the right purpose by the right people?

IndexTerms **– Big Data, U.S Cellular, Volume, Variety, Veracity, Velocity and Value**

Huge collection of data on internet refers to Big Data. In an era where everyone owns a smart phone and they are so much technically equipped, data is getting accumulated rapidly. The main features of big data are: Volume, Variety, Veracity, Velocity and Value. Volume indicates the sheer bulk of data generated every second in any application. It is shocking fact that every day the amount of data produced is nearing to 40-45 Zettabytes. Velocity talks about the speed at which these data are generated. The truth is, it's massive!! Researchers have shown that every 60 seconds, there are about 11 million instant messages, 1 million tweets and 180 millions of emails sent. Data is inconsistent in nature. Due to this, there might be some uncertainty in the data fields. Veracity shows this dubiety. Variety stipulates the different variants of data as indicated earlier to the discussion. Any organization considers the data only if it is of any value to their business. So the last V, Value deliberates this sense of data indicating whether the data is useful for the organization or if it is useless.

Any big data must first go through ETL stage where the data is Extracted, Transformed into the required format and Loaded for the further analysis. Once ETL stage is over, the interaction based on queries takes place, and based on the result of the queries, analysis through predictive study is made [1]. Today many companies are utilizing the benefits of the big data. Some of the examples are Google, Intel, and shopping websites such as Amazon, Flipkart, and many more [3].

Big Data Challenges

- Data Privacy and Safety: When the user's data goes public, one major factor that comes into picture is the security. We often know that when a user browse for any data, he actually leaves a digital trace online and easily his search pattern can be used to show some related ads in the next search time. So in such scenarios how the security and privacy is preserved is a big question. This has been one of the important research concerns for many researchers lately.
- Data Fraudulent: Information is everything now. With just a small piece of information, organizations are making millions of dollars today. And many cases, the data ends up into wrong hands. They have the bank account number, phone number and many important credentials of a user and may easily hack into user's information and misuse them.
- Data Compartmentalization: Managing this huge amount of data is really strenuous. There are various concepts of data warehousing to ensure the data is managed properly. However, a major catch is adopting the suitable algorithms for the purpose.

A small survey was conducted using U.S Cellular. They help to calculate the amount of data used by users' for various activities such as emails sent/ received, number of downloads, web pages visited and so on. The usage can be calculated on day or weekly basis. The survey included 10 people. Some of them are active on social media such as sharing pictures, videos and posts, some of researchers where they might have to download lots of stuffs related to research and few were average internet users. The study showed the following implications.

Table 1 – Survey Result of Data Usage by different internet users

Person Type	Data Usage (Per day)	Number of People considered for Survey and Data Usage
Researchers'	21 GB	4*21GB= 84GB
Social media and Shopping websites users	25GB	4*25GB=100GB

Normal Users	10 GB	2*10GB=20GB

A sample screenshot of one user involved in the survey is shown in figure.1. The slide bar helps the users to drag the number of entries required for each entity. On the left hand side of the panel, the red bar shows the amount of data used by that particular user. Experiments were conducted using the data of all the 10 users and the average amount of data usage is calculated for each person type and it is shown in the table 1.

Figure 1 – Sample screen shot of U.S Cellular that was used for the survey.

The survey results as indicated in the table 1 shows that on an average, only with 10 people, the amount of data generated per day is equal to 204GB. This is a huge data with just 10 people and when we calculate for the entire nation or the world, the amount of data generated looks frightening. All these data are available online and the organizations use them for their benefits. It is important to know that the study can be extended to multiple users and the results will vary accordingly. And after collecting this data, a researcher can work on the big data analysis. With this analysis, the organizations run their business and work on their strategies to impress the customers [2]. So next time, when they are online, trivially they might want to re-visit those websites.

To conclude, the amount of data generated is growing every second in a lightning speed and the use of this data is really important. Many organizations that are built around big data must take into consideration the challenges and should undoubtedly use the data ethically and merely for their business needs. Data has evolved to take so many new forms and accordingly the infrastructure and technologies to adopt this data must also be efficient and scaled-up to the demands. Further critical implications of big data must be addressed by the researchers and must be made known to all. This will educate a normal user as well and can avoid the upcoming evil consequences of this era where data is actually Very Big.

References:

[1] Jafar Raza Alam et al, "A Review on the Role of Big Data in Business", International Journal of Computer Science and Mobile Computing, Vol.3 Issue.4, April- 2014, pg. 446-453

[2] Rob Kitchin, "Small Data is the era of Big Data", GeoJournal Published on October 2014, Springer Publications, DOI: 10.1007/s10708-014-9601-7

[3] Sawyer (2008), "Data wealth, Data Poverty, Science and Cyberinfrastructure", Critical Studies in Innovation, 26(4), 355-371

Web References:

1. https://www.uscellular.com/data/data-estimator.html
2. https://www.flydata.com/the-6-challenges-of-big-data-integration/
3. https://en.wikipedia.org/wiki/Big_data
4. https://www.sas.com/en_us/insights/big-data/what-is-big-data.html