



ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2016; 2(2): 314-317
www.allresearchjournal.com
Received: 19-12-2015
Accepted: 21-01-2016

G Yasasvi
G D Goenka World Institute,
Lancaster University, England

Review paper on big data analytics

G Yasasvi

Abstract

In today's world information technology is booming and continues to have a big impact in the future. With number of internet and computing device users increasing, the information technology is going through a new phase where enormous amount of data is being generated this is called Big Data Analytics. In 2014 the number of Google search queries per minute was more than 4 million, in the same year Facebook collected 500 terabytes per day, in that there was a contribution of 300 million photos. By the end of year, photos and videos were summed up to 100 petabytes. The statistics goes on. These data are nothing but just a big bunch of information which are not yet sorted, or it can be said that the data is in raw form and needs to be processed. The term Big Data means that data generated is not only in huge amounts but is also of great variety and high rate velocity. Big Data has wide variety of applications varying from sector to sector where it can be used. For many multinational companies, Big Data is a source of revenue generation, for others it may can be a medium for carrying out mass feedback, such as for governments it can be used to know about the welfare of people and many more. Big Data is an evolving field of future interest with a great potential in it. Wherein information technology is rapidly changing with more data being generated, people who assess with the help of these data need to gain valuable acumens from it. Most of the data is being generated from social networks, internet browsing, online feedback systems and many more. This paper focuses on how the Big Data Analytics came into existence and evolved, its advantages, flip sides and future opportunities [1, 4].

Keywords: big data analytics, internet and computing device, high rate velocity

1. Introduction

With growing technology especially in the field of information, we can make coarse estimation about how much data is generated, not in terabytes but in zettabytes. The source for such huge amount of data are many, such as social feeds, internet, websites, online reviews, customer names and their choices for shopping or movies they like and many more. Companies thrive on these data and it's a big business for them by utilizing the data for the services that are customer oriented. But what happens if the data generated is not stored, if there are no such data storages? In such case, the organizations and Big Data firms would not be able extract crucial information or knowledge that could useful. Data is becoming a day to day commodity for the continuity of services for almost every organization [1, 2, 4].

There is an exponential rise of the data, with every second more and more data being generated. Increase in data storages, and capabilities and methods for data collection, the data is now accessible and available. Analyzing these data has also become easy as well. Storing data has also become more affordable due to availability of various platforms, thus having tough competitions among organizations for the extraction of high value information for further uses. When we talk about any data, the first thing that comes into the mind is whether it structured or unstructured [5].

There are basically three types of data generated, structured, semi-structured and unstructured.

Correspondence
G Yasasvi
G D Goenka World Institute,
Lancaster University, England



Source: of Big Data [16]

Structured Data

It refers to the data which exist in a fixed field in a file or database. It maybe the data, stored in the spreadsheets. It basically is predefined, what field of data is to be stored and how it is to be stored. The data which is predefined has a pattern which makes it easy to search. Some examples for structured data are digital library catalogue, census records, economic data, phone numbers, cargo and logistics manifest etc.

Semi structured Data

Semi-structured data is similar to structured data, in semi-structured data there is no formal structure associated with the relational databases or any other forms of data storage. But such data contains some types of tags or entity title that helps in distinguishing and separating them into categories (i.e. structured or semi-structured). Some of the examples for semi-structured are CSV files, XML, JSON, PDF documents etc.

Unstructured Data

It is a data that has no pre-definition or is not organized in an orderly manner. In general, these are the data sets that needs to be processed before any further use for decision making or any analysis. Few examples for unstructured data are web pages, multimedia content, email messages, word processing documents, audio files, natural voice languages, images, search engine queries, online streaming data etc.

Big Data Analytics

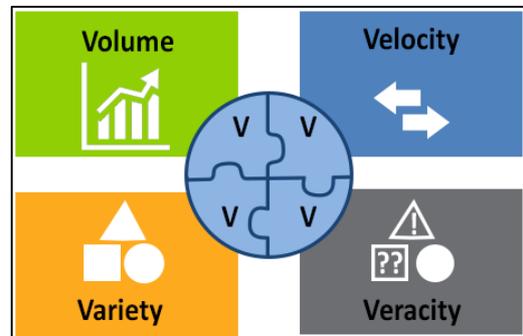
In recent times the term “Big Data” is being used for large datasets that has been growing exponentially, with huge data being generated every day, it’s difficult to work with using outdated databases for data analysis. From the term “Big Data” it could be understood that data is huge in size, variation, variety and veracity as well, which is not easy to store and manage simultaneously. The unstructured data is called as Big Data, this unstructured data is generated from everywhere and every day, from the internet, sensors, computing devices, phones, tablets etc. they are not in a

distinguishable manner and for the same reason it is called unstructured.

It is the process of examining large and varied data sets to uncover hidden patterns, correlations, trends, preferences, sentiment values and some other useful information that may help the organization in increasing their businesses. The data that is stored in databases are in the range of petabytes or more, which may further lead to some problems- capturing, storing, search, sharing, visualization and analytics. Thus, working on the problems help data analyst and scientist in increasing their accuracy in decision making by improving them [5, 6]. With exploration of huge volumes of data, it has become a common practice for many information technology companies to compete in the market [13, 14, 16].

Characteristics of Big Data

Big Data is very large, distributed and diverse to work with. Majorly classified into 4 V’s, volume, Variety, Velocity and Veracity. The following figure shows the aspects to be considered while working on Big Data [8].



Characteristics of Big Data [10]

Volume

Volume is a self-explanatory word, with the word “Big” in Big Data, it obviously means that the volume is enormous. The data storage units for “Big Data” are not in Gigabytes or Terabytes but in Petabytes. With the exponential rise in the data generation, the information or data collected and stored is huge [8, 9]. These multiple sources are social media, online review portals, customer details, healthcare, entertainment, government and many more. Such amounts of data require big storage devices or servers such as distributed systems or cloud storages.

Distributed systems are a network of multiple computers or systems that are connected to each other via middleware to share each other’s resources such processing power, memory storage, be accessible whenever needed. Ex – Hadoop, Apache, Google BigQuery, HPC etc.

Cloud storage is a model for data storage where the digital data is stored, the physical storage spans multiple servers. It used for data storage and access whenever required from any location. Ex- Amazon Web Service (AWS), Microsoft Azure, Google Drive etc.

Variety

Since the data collected from various websites and devices across the globe, it is very difficult to understand which part belongs to what. Thus, by variety we can understand and differentiate certain data based upon the format it has, the types of data, the uses and the ways of analyzing it.

Velocity

It deals with the rate at which data flows in from various sources such as internet. This flow is massive and continuous. The flow of real time data helps data scientist and business analyst in making valuable decisions that may provide them with a strategic competitive lead over their opponents [8].

Veracity

Veracity means that if the data can be trusted, or if the data is reliable source for decision making. Every data analyst or scientist knows that there are discrepancies which are inherited while collecting the data [8, 9]. While carrying out data analysis or business decision making one should keep in mind about of veracity as well.

Application areas

Targeting customers to build clientele

Attracting customers and targeting them is one of the biggest areas where the Big Data is used. In this sector, the aim of any business is to understand about the target market and the customers falling in that target segment. The data that the companies collect from online review portals, social media, browsing internet, feedbacks etc. are the major source for such companies. With data extraction from all possible ways, it can be used to understand the customers and their needs, and accordingly cater to them [12, 14, 15].

By analyzing the generated data and using predictive modeling, they try to target the customers. For example, in case of a make-up manufacturer, who wants to launch a new make-up product, then before the launch the company will try to locate the possible targets for their product through online and offline surveys. Based upon the results received after the analysis, the company makes decision about the launch of the product.

Big Data Analytics is not only for big corporations and governments, it is also useful for common people as well. People who like to research about a product or service can decide whether to go ahead with buying the service or not, which in turn, is very beneficial. For example, if someone wants to buy a new coffee machine and is confused between different brands for the machine, then he/she can simply check the reviews for the machine given by other customers. Based upon the rating and the reviews given, the customer can decide whether to buy or not.

Supply Chain Management

Areas such as supply chain management are very important for any business to run. These areas can be more optimized by using Big Data analytic, for more efficient and optimized business. If a small retailer for example, has a store of confectionary and food items, based upon the market research and the predictions available online, the retailer can stock up his warehouse based upon future demand for a certain product. In this way the retailer can be efficient in running his/her business without effecting any other factors [3, 15].

Healthcare

Big Data techniques for healthcare are changing and are effective after the implementation. They are being used to monitor infants and patients to monitor their pulse rate and heartbeat. Not only this, these days, blood pressure counts, sleeping patterns etc. are being monitored and these data are

further used for patient health analysis. By recording each minute of breathing and pulse, medical practitioner and doctors can predict diseases 24 to 36 hours early. Such predictions help doctors in better treatment of small infants and adult patients [15].

Government

Big data gives tools and other important information to government that helps them identifying poverty ridden areas, and assess and implement new and effective ideas to reduce poverty. Big data is very useful tool for government to help people, and can solve many problems, but if it's in right direction.

Education

Education is one of the main priorities for any government. To improve education, there are a lot of different things to be considered, for example up-to-date data, relevant information etc. With valid and vital data, governments can assess the local and state level needs of education and ensure that good education is served to its citizens.

Advantages

After the emergence of Big Data Analytics, people and big corporations are getting benefits from it. Below are the few advantages of using Big Data Analytics.

Distributed data storage – tools such as Hadoop, can store large amounts of distributed data and helps in increasing the efficiency of the business by keeping them intact and ready whenever required.

Time reduction – with new tools being introduced for Big Data less time is required for data analytics, thus reducing the time in decision makings.

New product development – if a company knows the current market trend or the upcoming trend based on the customer's needs, a company can decide what product to launch and when to launch. Thus, giving a strategic position in the market.

Optimization - for business which are supply chain based, by using Big Data Analytics, can improve and optimize their business with more efficient results.

Quality Management – with growing business in telecom, manufacturing and food industry, big data analytics can be used for quality management and improvement.

Limitations

Since, a coin has two faces, Big Data also has disadvantages as well.

Storage - with increase in data generation from various sources, it becomes difficult to store that data. Therefore, storage will be another big issue in the future.

Processing capabilities - with growing demand from big corporations for data analytics, more data analysis, there is a need for high computation and processing powers.

Visualization - it is the representation of the data in a systematic form. Big Data in high in volume, variety and

velocity, so only good and vital information should be processed and visualized.

Inconsistency - if a company is extracting data from certain site, the data should be consistent rather changing day to day. No consistency means that there is no proper data analysis on the data collected thus leading to unprecise and inaccurate results.

Security - as other technological endeavors, Big Data Analytics is also prone to data breach. The information can be leaked to competitors or can be made public for different reasons.

Transferability - since most of the data lies around the internet, it is important to know how efficiently can the data be transferred from public server to a private server or private cloud without harming the firewall. It should be easy for the specialist also to easily transfer the data for repeat analysis

Conclusion

The aim of this paper was to highlight about Big Data Analytics. An area of interest which has gained a high value because of its imminent values and great opportunities to explore. Also, application areas of Big Data Analytics, advantages and limitations associated to it. Most importantly its characteristics, because it defines the actual need of Big Data Analytics. If utilized and implemented correctly, Big Data can bring out several potential and new innovations in the world of rapidly changing information technology. In coming years, Big data has a great significance, with the coming era being data driven, it will provide an unforeseen insight and also benefits for business analyst in decision making.

References

1. Chong D, Shi H. Big Data analytics: a literature review. *Journal of Management Analytics*. 2015; 2(3): 175-201.
2. Elgendy N, Elragal A. Big Data Analytics: A Literature Review Paper, Department of Business Informatics and Operations, German University (GUC), Cairo, Egypt, 2014.
3. Big Data analytics: What it is and why it matters, *Sas.com*. [Online]. Available: https://www.sas.com/en_us/insights/analytics/big-data-analytics.html#. Accessed: 2015.
4. Data Analytics Trends that Will Make Waves in 2014 Transforming Data with Intelligence", *Transforming Data with Intelligence*. [Online]. Available: <https://tdwi.org/Articles/2014/01/28/5-Data-Analytics-Trends-2014.aspx?Page=2>. [Accessed: 12- Nov- 2015].
5. Big Data: 20 Free Big Data Sources Everyone Should Know", *SmartData Collective*. [Online]. Available: <https://www.smartdatacollective.com/big-data-20-free-big-data-sources-everyone-should-know/>. [Accessed: 06- Nov- 2015].
6. "The 4 V's of Big Data - dummies", *dummies*. [Online]. Available: <http://www.dummies.com/careers/find-a-job/the-4-vs-of-big-data/>. [Accessed: 09- Nov- 2015].
7. "The Four Vs of Big Data", *IBM Big Data & Analytics Hub*. [Online]. Available:

- <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>. [Accessed: 2015].
8. "Four 'V's of Big Data: volume velocity variety veracity", *TheServerSide.com*. [Online]. Available: <http://www.theserverside.com/feature/Handling-the-four-Vs-of-big-data-volume-velocity-variety-and-veracity>. [Accessed: 04- Dec- 2015].
9. "The 4 V's of Big Data - Zarantech", *Zarantech*. [Online]. Available: <http://www.zarantech.com/blog/the-4-vs-of-big-data/>. [Accessed: 04- Nov- 2015].
10. Cohen J, Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C.: MAD Skills: New Analysis
11. Practices for Big Data. Proceedings of the ACM VLDB Endowment 2(2), 1481–1492 v (2009). Casoria, A., Song, I., Davis, K.C.: Analytics over Large-Scale Multidimensional Data:
12. The Big Data Revolution! In: Proceedings of the ACM International Workshop on Data Warehousing and OLAP. 2011, 101-104.
13. Adams MN. Perspectives on Data Mining. *International Journal of Market Research*. 2010; 52(1):11-19.
14. Cebr: Data equity, Unlocking the value of big data. in: SAS Reports. 2012, 1-44.
15. "How is Big Data Used in Practice? 10 Use Cases Everyone Must Read", *Bernard Marr*. [Online]. Available: <https://www.bernardmarr.com/default.asp?contentID=1076>. Accessed: 08- Nov- 2015.
16. S. Bansal and S. Kagemann, "Integrating Big Data: A Semantic Extract-Transform-Load Framework, *Computer*. 2015; 48(3):42-50.