An Analysis of Big Data Analytics

Shri Ram College

Shweta Jain, Jains812@gmail.com

**Abstract:** The great bulk of the info in our world is available on demand. While watching a video stream, accessing social media, playing games, or utilizing GPS to find a place, people and devices are continually producing data. This data is generated from numerous sources using a wide range of methods and technologies, and it grows every day. The information is classified as "Big Data". Big Data is enormous in terms of variety, speed, and sheer volume. It is diverse in nature and contains both organized and unstructured data. Big Data analysis seeks to identify important values, make recommendations, and/or provide decision support. We present a thorough overview of the big data analytics research in this topic while emphasizing a particular area of concern. We examine six different forms of big data applications, including structured data analytics, text analytics, web analytics, multimedia analytics, and mobile analytics, in accordance with Application Evolution. We provide examples of large data analysis methods such as classification, crowd sourcing, A/B testing, and data mining.

 Keywords: Big data management, analytics, and technique for analysis are some related terms.

1. Introduction

The term "big data" was first used in a Silicon Graphics (SGI) PowerPoint presentation in 1998. John Massey wrote a book titled Big Data [3]. Big data is primarily very huge and complex in size. At all stages of the process that might extract value from data, big data challenges with heterogeneity, scalability, timeliness, complexity, and privacy obstruct development [5]. Email attachments, social media posts, audio and video information, and multiple database tables are just a few examples of the sources of big data. People use Twitter in a variety of ways, and 250 million tweets are archived every day. 4 billion people watch YouTube every day. Today, data is produced in zettabytes. Among the many sectors where big data is used, there are financial services, healthcare, retail, web/social, manufacturing, and government. can be used effectively [10]. Big data is becoming a part of every sector of the global economy. We predict that each company in the US economy with more than 1,000 employees has an average of 200 terabytes of data stored by the year 2005 [12]. Big data is still evolving swiftly because the underlying technology is improving. In August 2010, the OMB at the White House proclaimed big data to be a national priority alongside healthcare and national security.

Traditional data management and analysis solutions are built on relational database management systems (RDBMS). RDBMS and Big Data differ in the following two areas:

1. RDBMS only supports structured data, whereas big data supports semi-structured and unstructured data.
2. RDBMS cannot scale up to deal with massive amounts of data. privacy hardware. They also cannot connect with commodity hardware in parallel.

Analytics transitions to big data analytics when? The scope of big data has expanded. The management of the millions of data points included in US census information alarmed delegates of the inaugural VLDB (Very large databases) conferences in 1975 [8]. Large datasets with a range of data types are examined using big data analytics.

For example, unidentified relationships, market trends, consumer preferences, and other relevant information [16]. Better marketing and customer service may result from the analytics.

The use of big data analytics projects is quickly becoming the go-to response to business and technological changes that are upending conventional data management procedures [10]. Analytics aids in identifying changes and potential fixes. [5].Using big data analytics, the user is attempting to unearth previously undiscovered business truths.

Enterprise was aware earlier [7]. In part 2, we present a literature review of big data analytics. Background and an overview of big data are provided in Section 3. Big data analytics are covered in detail in Section 4, and the study is wrapped up in Section 5.

**2. Literature survey**

Over the course of several years, numerous academics have successfully finished their work on big data. Numerous pieces have been published in the mainstream business media, including Forbes, Fortune, Bloomberg, Business Week, The Wall Street Journal, and The Economist [1]. High data volume, high data velocity, and high data representation capabilities are characteristics of big data, according to the National Institute of Standards and Technology [NIST] [15]. In order to launch a big data research plan, the Obama Administration announced in March 2012 that the US would invest $200 million [2].

From 130 Exabyte’s to 40,000 Exabyte’s, or a doubling every two years, the global data volume would expand by a factor of 300 between 2005 and 2020, according to an IDC Report [9]. IBM claims that 2.5 quintillion. 90% of which was produced in the previous two years. Social networking services like Facebook, LinkedIn, and Twitter are estimated to have 750 million, 110 million, and 250 million users, respectively [17]. Big Data from business, government, and the research community has resulted in an emergent a field of study that has generated a lot of attention. The public media's and industry reports' coverage, for instance, is the first illustration of the widespread interest: New York Times, The Economist [12]. The best way to collect information about people from all angles is through mobile phones, and the vast amounts of data that mobile carriers can analyze have a positive impact on our daily lives [13]. Figure 1 shows a graph that suggests the amount of data practically grew starting in 2005. However, take into account the exponential expansion in data since the year 2005, when user and corporate system level data began to pour into the data warehouse [11].



**Figure 1:** Exponential growth of data from year 2005 to 2012[11]

when the Data Warehouse's capacity increased from 50 GB to 1 TB or 100 TB. When data was created from numerous organizations, it was in a structured format. Three characteristics—volume, variety, and velocity—guide the flow of data. Many businesses were struggling with the issue of how to increase the data warehouse's capacity to accommodate the increased requirement.

By comparing the types of data generated and saved, such as whether the data is in audio, video, image, or text format, Figure 2 indicates that there are variances in the amount of data stored in various sectors [12]. The sectors in charge of text and numeric data are banking, insurance, and healthcare. Audio and video data are largely the result of communication and media.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Sectors* | *Video* | *Image* | *Audio* | *Text/Numeric* |
| Banking |  |  |  |  |
| Insurance |  |  |  |  |
| Retail |  |  |  |  |
| Wholesale |  |  |  |  |
| Utilities |  |  |  |  |
| Health care |  |  |  |  |
| Transportation |  |  |  |  |
| Communication & Media |  |  |  |  |
| Construction |  |  |  |  |
| Government |  |  |  |  |
| Education |  |  |  |  |

Penetration:



Figure 2: Data generation and growth might vary depending on the sector by utilizing formats like audio, video, etc. [12].

# Big Data

Big data is a new phrase for extremely massive and intricate datasets. Without new technology, managing huge datasets is challenging. A paper on big data that details the numerous business prospects that big data presents was produced by the Mckinsey Global Institute (MGI) [12]. One of the writers, Paulo Boldi, states that Big Data requires big intelligence, not huge machines, according to [6]. Following are the two categories of big data:

**Data That Is Structured**

 These data are simple to analyse. It is presented as numbers, figures, transaction data, etc.

**Data That** **Isn't Structured**

These data include complicated information like email attachments, social networking site remarks, and images. These data are difficult to analyse.

The three v's of big data management were introduced by Doug Lancy [3]:

**a.Volume**: This term refers to the quantity of data. It alludes to enormous amounts of data.

b.**Variety:** It discusses many data and source kinds, such as structured, semi-structured, and un structured data.

**C. Velocity** - This term describes how data is moving. Quickly generated, processed, and analyzed data.



**Figure 3:** 3v’s Big Data management

1. **Big Data Analytics**

Organizations can use big data analytics to evaluate a variety of structured, semi-structured, and unstructured data in quest of insightful business insights. In June 2011, the Mckinsey Global Institute, a think tank inside Makinsey, released a significant report on big data [12]. Big Data is "a key basis of competition and growth," according to its overwhelming conclusion. Any decision-making based on data is frequently referred to as analytics (including its Big Data variant) [8]. Corporate analytics and Academic research analytics are two subcategories of the term analytics. Team applies their knowledge of statistics and data mining to corporate analytics. Researchers examine data in academic analytics to test hypotheses and develop theories [8].

Researchers discovered that the generated data in big data analytics was shared among several big data applications, including the ones listed below [2].

**Detailed Analyses**

For structured analytics, a lot of data is generated from the commercial and scholarly research areas. These data are managed utilising RDBMS, data warehousing, OLAP, and BPM. Numerous research areas, such as e-commerce and privacy-preserving data mining, enhanced the amount of data.

**Analytics for text**

Documents, emails, and social media posts all contain text, which is one of the most used data storage formats in text analytics. The process of extracting important information from a large body of text is known as text analytics, sometimes known as text mining. The cornerstones of text mining systems are text representation and Natural Language Processing (NLP), with an emphasis on the latter. [2].

**Google Analytics**

Retrieving and extracting data from web pages is the goal of web analytics. Web mining is another name for web analytics.

**Analytics in Multimedia**

Recent years have seen a sharp rise in the amount of multimedia data, including images, audio, and video. The procedure of separating intriguing data and semantics from multimedia data is referred to as "multimedia analytics." Audio summarization, multimedia annotation, and multimedia indexing and retrieval are only a few of the many areas that are addressed by multimedia analytics.

**Cellular Analytics**

At the end of 2012, mobile data traffic climbed by 885 PBs each month. Mobile analytics are made possible by a large volume of applications and data. RFID, mobile devices, sensors, and more are all part of mobile analytic.

**5. Big Data Analysis Methodology**

Dataset analysis can be done using a variety of methods. Machine learning is used in some methods. Analyze new dataset combinations using these techniques [12].

**Testing A/B**

a method where a control group is compared to multiple test groups to ascertain whether alterations may enhance a certain variable, such as the marketing response rate.

**Classification**

A method for classifying fresh datasets into established categories and assigning to them, for instance: Whether a mushroom is dangerous or edible[4]. It's employed in data mining.

**Using the crowd to source**

a method of gathering data from a huge community or group of people, called the "crowd." Typically, it happens via network media, like the web.

**Data Analysis**

a method for extracting data patterns from huge datasets of statistical and machine learning combinations.

 **6. Conclusion**

The concept of big data has been covered in this essay. Big data refers to vast, complex datasets that are created from a variety of sources, like playing video games, posting comments on social media, downloading email attachments, etc. As seen by example A6, big data is complex.

The concept of big data has been covered in this essay. Big data refers to vast, intricate datasets that are created from a variety of sources, including social media comments, playing video games, email attachments, etc. Big data is complicated due to its amount, variety, and velocity. Big data analytics are more difficult to apply to these three concepts. The literature revie Creating and storing data in any format, including text, audio, video, and photos, is feasible. Researchers have separated the generated data for big data analytics into a variety of applications, including structured data analytics, text analytics, online analytics, multimedia analytics, and mobile analytics. More research is required to address many large data system issues. The velocity, variety, and volume of big data research can increase government sector efficiency and profit for enterprises. Big data analytics are more difficult to apply to these three concepts. The literature review we've presented demonstrates the industry-wide exponential rise of data starting in 2005. Whether the data is music, video, photos, or text, there are different ways it can be created and stored. Researchers separated the gathered data in big data analytics we've presented demonstrates the industry-wide exponential rise of data starting in 2005. There are differences It is possible to generate and store data in audio and video formats for use using text analytics, structured data analytics, online analytics, multimedia analytics, and mobile analytics, among other big data applications. To address a number of huge data system challenges, more study is necessary. Common big data applications can help businesses and increase efficiency in the public sector.

**7. Acknowledgement**

I want to express my gratitude to everyone who helped me thoroughly prepare this paper. I would also like to thank my guide for helping me and giving me the right advice. Finally, I would want to thank every website and journal article that I used as references to effectively complete my review paper.

1. **References**

[1] Sameera Siddiqui, Deepa Gupta,” Big Data Process and Analytics: A Survey”, International Journal Of Emerging Research in Management & Technology, ISSN: 2278-9359, Volume 3, Issue 7, July 2014.

[2] Han Hu, Yongyang Nen, Tat Seng Chua, Xuelong Li,” Towards Scalable System for Big Data Analytics: A Technology Tutorial”, IEEE Access, Volume 2, Page No 653, June 2014.

[3] Bharti Thakur, Manish Mann,” Data mining for big data: A Review”, International journal of advanced Research in Computer Science and Software Engineering, ISSN: 2277 128x, Volume 4, Issue 5, May 2014.

[4] Anand V. Saurkar, Vaibhav Bhujade, Priti Bhagat and Amit Khaparde,” A Review Paper on Various Data Mining Techniques”, International Journal of Advanced Research in Computer Science and softwareEngineering, ISSN:2277 128X,Volume 4, Issue 4, April 2014.

[5] Puneet Singh Duggal, Sanchita Paul, ”Big Data Analysis: Challenges and Solutions”, International Conference On Cloud, Big Data and Trust 2013, Nov 2013.

[6] Albert Bifet, “Mining Big Data in Real Time”, informatica, 2013.

[7] Stephen Kaisler, Frank Armour, J. Alberto Espinosa and William Money,” Big Data: Issues and Challenges Moving Forward”, Hawaii International Conference on System Science, IEEE Computer Society, Page No. 995, 2013.

[8] D.Fisher, R.Deline, M.Czerwinski and S. Drucker,”Interaction with big data analytics”, Volume 19, No.3, May 2012.

[9] J.Gantz, D. Reinset,” The Digital Universe in 2020: Big Data, Bigger digital shadow, and biggest growth in the far east”, in Proc : IDC iview, IDC Anal, Future, 2012.

[10] Denis Guyadeen , Rob Peglar,” Introduction to Analytics and Big data- Hadoop”, SNIA Education Committee, 2012.

[11] Neil Raden,”Big Data Analytics Architecture”, Hired Brains Inc, 2012

[12] James Manyika, Michael Chui, Brad Brown, Jacques Bhuhin, Richard Dobbs, Charles Roxburgh, Angela Hungh Byers, “Big Data: The next frontier for innovation, competition and productivity”, June 2011.

[13] Wei Fan, Albert Bifet, “Mining Big Data: Current Status and Forecast to the Future”, SIGKDD Explorations, Volume 14, Issue 2.

[14] American Institute Of Physics(AIP), 2010. College Park, MD(http:// www.aip.org /fyi/2010/)

[15] M.Cooper, P.Mell(2012). Tackling big Data(Online). Http://csrc.nist.gov/groups/SMA/Forum/document/June2 012Presentation/f%CSM\_june2012\_cooper\_Neul.pdf.

[16] www. Searchbusiness analytics.techtarget.com

[17] www.ebizmba .com/articles/social-networking-websites.