International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

A Review Paper on Big Data Analytics

Ankita S. Tiwarkhede[1], Prof. Vinit Kakde[2]

[1]ME (CSE) 2[ND] Sem, GHRCEM, SGBAU Amravati University, Amravati, India

[2]GHRCEM, Amravati, India, SGBAU Amravati University

Abstract: ***We live in on-demand world with vast majority of data. People and devices are constantly generating data, while streaming a video, active in social media, playing games, search any location using GPS. This data increase day by day from many resources, various types of techniques and technologies. The data is categories as "Big Data". Big Data is huge in Variety, Velocity and Sheer volume. It is structured and unstructured data and heterogeneous in nature. The goal of Big Data analysis is to extract useful values, suggest conclusions and/or support decision making. In this topic, we provide an extensive survey of big data analytics research, while highlighting the specific concern in big data world. According to Application evolution, we discuss six types of big data application such as structured data analytics, Text analytics, Web analytics, Multimedia analytics, and Mobile analytics. We illustrate the techniques of analyzing the big data such as A/B testing, classification, crowdsourcing, and data mining.***

Keywords: Big data management, Big data, Analytics, Analyzing Technique

1. Introduction

Big Data term appeared for First time in 1998 in Silicon Graphics (SGI) Slide Deck By john Massey with the title of Big Data [3]. Big Data is very vast in majority and Complex data. Heterogeneity, scale, timeliness, complexity, and privacy problems with big data hamper the progress at all phases of the process that can create value from data [5]. There are various resources of Big Data For Example: Audio, Videos, and Post in Social Media, Various Database Tables, and Email Attachment etc. People uses twitter in diverse form and store 250 Million tweets Per Day. 4 Billion People watching YouTube per Day. Nowadays, Data produced in Zettabytes. Big data has many opportunities like financial services, Healthcare, Retail, Web/social, Manufacturing and Government [10]. Big data has now reached every sector in the global economy. We estimate that by 2005, nearly all sectors in the US economy had an average of 200 terabytes of stored data per company with more than 1,000 employees [12]. Big data moving continue to evolve rapidly, driven by innovation in underlying technologies. In August 2010, The White house, OMB, Proclaimed that big data is national challenge and priority along with healthcare and national security [14].

Traditional data management and analysis system mainly based on Relational database management system (RDBMS). There are two aspects in which RDBMS and Big Data differs:

1) RDBMS can support structured data but big data supports for semi-structured and unstructured data.
2) RDBMS scale up to expensive hardware and cannot connect with commodity hardware in parallel and it's not supported by big data.

When does analytics become big data analytics? The size that defines big data has grown. In 1975 attendees of the first VLDB (Very large databases) conferences worried about handling the Millions of data points found in US census Information [8]. Big data analytics is the process of examining large datasets containing a variety of data type's i.e. Unknown correlations, market trends, customer preferences and other useful information's [16]. The analytics can more lead to more effective marketing, better customer services.

Big data analytics project are rapidly emerging as the preferred solution to address business and technology trends that are disrupting traditional data management processing [10]. Analytics helps to discover what has changed and the possible solutions [5].With big data analytics, the user is trying to discover new business facts that no one in the

Enterprise knew before [7]. We introduce literature survey of big data analytics in section 2. Section 3 contains background and overview of big data. Section 4 contains big data analytics in detail and section 5 concludes the paper.

2. Literature survey

Over the last many years, there are many researchers has completed their work successfully on big data. Hundreds of articles have appeared in the general business press (For example Forbes, Fortune, Bloomberg, Business week, The Wall street journal, The Economist)[1]. National Institute of Standards and Technology [NIST] said that Big Data in which data volume, velocity and data representation ability to perform effective analysis using traditional relational approaches [15]. In March 2012, The Obama Administration announced that the US would invest 200 Million Dollars to launch a big data research plan [2].

An IDC Reports predicts that from 2005 to 2020, the global data volume will grow by a factor of 300, from 130 Exabyte's to 40,000 Exabyte's, representing a double growth every two years[9]. IBM estimates that everyday 2.5 quintillion bytes of data are created out of which 90% of the data in the world today has created in the last two years. It is observed that social networking sites like Facebook have 750 Million users, LinkedIn has 110 million users and Twitter has 250 million users [17]. From industry, government and research community, Big Data has led to an emerging

research field that has attracted tremendous interest. The broad interest is first exampled by coverage on both industrial reports and public media for example: The economist, New York Times [12]. Mobile Phones becoming best way to get data on people from different aspect, the huge amount of data that mobile carrier can process to improve our daily life [13]. In figure 1, From Year 2005, it would appear from this graph that the amount of data was practically increased. However, Consider exponential growth in data from 2005 year, when enterprise system and user level data was flooding into data warehouse [11].

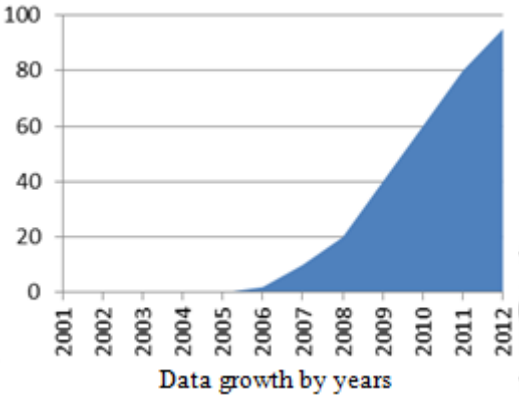

Figure 1: Exponential growth of data from year 2005 to 2012[11]

When the capacity of Data Warehouse grew from 50 GB to 1 TB – 100TB. Data was in structured form when it creates from many organizations. Data goes from three properties like volume, Variety and velocity. Many companies were facing the problem on how to expand the capacity of data warehouse to accept the new requirement.

Figure 2 illustrates that there are variations shows in the amount of data stored in different sectors by using the types of data generated and stored i.e. whether the data is in audio, video, images and text format and differ from industry to industry[12]. Banking, Insurance and Health care sectors are responsible for text/numeric data. Communication and Media are highly responsible for audio and video type of data.

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

Penetration: High Medium Low

Figure 2: Variations possible in generating and growth of data by using types such as audio, video etc.in various sectors [12].

3. Big Data

Big data is the new term that contains large and complex datasets. It is difficult to manage these datasets without new technology. The Mckinsey Global Institute (MGI) published a report on big data that describes the various business opportunities that big data opens [12]. Paulo Boldi, One of the authors says "Big Data does not need big machines, it needs big intelligence" [6]. There are two types of Big Data is as follows:

3.1 Structured Data

These data can be easily analyzed. It is in numerical form, figures, and transaction data etc.

3.2 Unstructured Data

These data contain complex information such as Email attachments, Images comments on social networking sites. These data cannot be easily analyzed.

Doug Lancy was the first one talking about 3v's in big data management [3]:
Volume - It describes the amount of data. It refers to mass quantities of data.
Variety - It describes different types of data and sources including structured, semi-structured and unstructured data.
Velocity - It defines the motion of data. Data created rapidly, processed and analyzed.

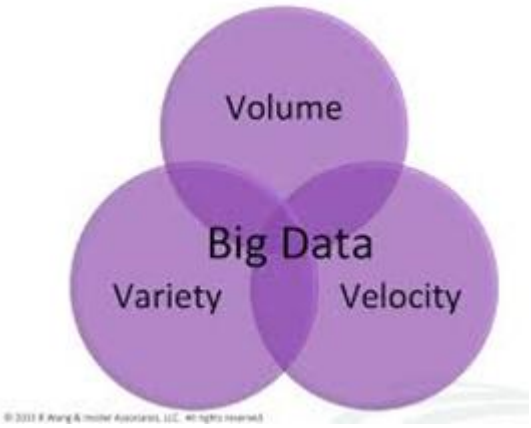

Figure 3: 3v's Big Data management

4. Big Data Analytics

Big data analytics enables organizations to analyze a mix of structured, semi structured and unstructured data in search of valuable business information. Makinsey's internal Think-Tank, the Mckinsey Global Institute, published a major study in June 2011 on Big Data [12]. Its overloading conclusion: Big Data is "a key basis of competition and growth". The term Analytics (including its Big Data form) is often used broadly to cover any data-driven decision making [8]. The term analytics divided into two groups: Corporate analytics and Academic research Analytics. In Corporate Analytics, Team uses their expertise in statistics and Data mining. In Academic Analytics, Researchers analyze data to test Hypotheses and form theories [8].

In Big Data Analytics, Researchers found that the generated data divided into various Big Data application such as follows [2].

4.1 Structured Analytics

In structured analytics, large quantity of data is generated from business and scientific research fields. These data is managed by RDBMS, Data warehousing, OLAP and BPM. Data grown by various research area like Privacy preserving data mining, E-commerce.

4.2 Text Analytics

In Text analytics, Text is one of the most common forms of storing the information and it includes Email communication, documents, and Social media contents. Text analytics also known as Text mining, refers to the process of extracting useful information from large text. Text mining system is based on text representation and Natural Language Processing (NLP) with emphasis on the latter [2].

4.3 Web Analytics

The aim of Web analytics is to retrieve, extract the information from Web Pages. Web Analytics also called Web mining.

4.4 Multimedia Analytics

Recently multimedia data, including images, audio, and video has grown at a tremendous rate. Multimedia analytics refers to extract interesting knowledge and semantics captured in multimedia data. Multimedia analytics covers many subjects like Audio Summarization, Multimedia annotation, Multimedia indexing and retrieval.

4.5 Mobile Analytics

Mobile data traffic increased 885PBs Per Month at the end of 2012. Vast volume of application and data leads to mobile analytics. Mobile analytics involves RFID, mobile phones, Sensors etc.

5. Technique for Analyzing Big Data

There are many techniques that can be used to analyze datasets. Some techniques are machine learning. From this techniques, analyze new combination of datasets [12].

5.1 A/B Testing

A technique in which control group compared with various test groups in order to determine what changes will improve a given variable for example- Reponse rate of marketing.

5.2 Classification

A technique in which to identify the categories of new datasets and assign into predefined classes for example- classification of mushroom as edible or poisonous[4]. It is used for data mining.

5.3 Crowd Sourcing

A technique in which collecting data submitted by large group of people or community i.e. crowd. It is usually through network media such as web.

5.4 Data Mining

A technique in which extracts patterns of data from large datasets of combinations from statistics and machine learning.

6. Conclusion

In this paper, we have presented the concept of big data. Big data is the large and complex datasets and it is generate from various sources like social media comments, playing a video game, email attachments etc. There is complexity in big data such as velocity, variety and volume. These three terms are more challenging for big data analytics. We have provided literature survey shows exponential growth of data in industries from 2005 year. There are variations possible while generating and storing data whether data is in audio, video, images and text. In big data analytics, Researchers divided generated data into various big data application such as structured data analytics, text analytics, web analytics, multimedia analytics and mobile analytics. Many challenges in the big data system need further research attention. Research on typical big data application can generate profit for businesses, improve efficiency of government sectors.

7. Acknowledgement

I would like to thank to all people who help me prepare this paper completely. I would also thank to my guide who help me and get proper suggestion. Finally I like to thank to all website and journal papers which I have refer to create my review paper successfully.

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 software

Engineering, 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/June2012Presentation/f%CSM_june2012_cooper_Neul.pdf.
[16] www. Searchbusiness analytics.techtarget.com
[17] www.ebizmba .com/articles/social-networking-websites.