Feature Analysis of Big Data Usage in Different Platform

R.Stalin, M.Sc., M.Phil.,
Assistant Professor, Department of Computer Science
Morappur Kongu College of Arts & Science
Morappur, Dharmapuri(Dt) Tamilnadu, India

OPEN ACCESS

Volume : 6

Special Issue : 1

Month : September

Year: 2018

ISSN: 2321-788X

Impact Factor: 3.025

Citation:
Stalin, R. (2018). Feature Analysis of Big Data Usage in Different Platform. *Shanlax International Journal of Arts, Science and Humanities*, 6(S1), pp.67–71.

DOI:
https://doi.org/10.5281/zenodo.1410985

Abstract

Big data is a huge amount of data set. It plays a main role in handling large and complex data where the traditional data processing application software is inadequate to deal with the big data. Big data focuses on capturing data, data storage, data analysis, search, sharing, transfer, visualization, query and updating information privacy. Relational database management systems and desktop statistics- and visualization-packages often have difficulty handling big data. The work may require "massively parallel software running on tens, hundreds, or even thousands of servers." What counts as "big data" varies depending on the capabilities of the users and their tools, and expanding capabilities make big data a moving target. The main characteristics of big data are volume, variety, velocity, machine learning and digital footprint. Nowadays, big data is applied to different domains like healthcare, commerce, cloud computing and so on. This paper presents an overview of big data and analyses the domains which adapts big data to store a large amount of data.

Keywords: Big data, cloud computing, commerce, healthcare

Introduction

Since the advent of the Internet and much more after social media introduction, the data has grown by leaps and bounds every day. The word 'BIG DATA' was first mentioned by Michael Cox and David Ellsworth [1]. The explosion of data social media, multimedia, data warehouses and the Internet of Things produce a huge amount of structured and unstructured data. Towards the investigation of these large volumes of data, big data is emerged and developed to manage a large amount of data.

Dissemination of the internet increases continually, as the number of people browses the web. People use mail and social media to communicate with others which generate a huge amount of data. Apart from that, a large amount of mobile data are generated day by day through the internet, sensor and mobile devices. Adding to that data analysis is the basis for the many fields like science, engineering, and management. It is essential for localized data or location data to optimize and personalize the local organization data, unlike web-based big data.

Big data is emerging and essential in the current scenario to manage and maintain the big amount of data. Big data is implementing in different domains like healthcare, commerce, engineering, science

and cloud platform. Many more algorithms are developed to optimize big-data and to assure the security of big data. This paper gives an overview of the features of big data and the challenges of using big data in a different domain.

The paper was organized as follows: Section 2 gives the brief notes on big data. Section 3 analyses the usage of big data in a different domain. Conclusions are proposed in section 4.

State-of-the-art of Big Data

Big Data is data that is too large, complex and dynamic for any conventional data tools to capture, store, manage and analyze. Traditional tools were designed with a scale in mind[2].

If the requirement of this Organization increases over time or if it wants to run more than granular analysis, it had to re-invest in data infrastructure. The cost of resources involved in scaling up the resources typically used to increase exponentially. Further, there would be a limitation on the size you could scale up to (e.g., size of machine, CPU, RAM, etc...). These traditional systems would not be able to support the scale required by some of the internet companies.

Fortunately or unfortunately, there is no size / parametric cut off to decide whether the data is "big data" or not. Big data is typically characterized[3] basis what is popularly known as 3 Vs.

- Volume – Today, there are organizations producing terabytes of data in a day. With increasing data, you will need to leave some data without analyzing, if you want to use traditional tools. As data size increases further, you will go away more and more data without analysis. Thus means leaving value on the table.
- Variety – While volume is just the start, selection is what makes it difficult for traditional tools. Traditional tools work best with structured data. They require data is a particular structure and format to make sense out of it. However, the flood of data coming from emails, customer comments, social media forums, customer journey on website and call-centers are unstructured or semi-structured at best.
- Velocity – The pace at which data gets generated is as critical as the other two factors. The speed with which a company can analyze data would eventually become a competitive advantage for them. It is their speed of analysis, which allows Google to predict the location of full patients' almost in real time. So, if you are unable to analyze data at speed faster than its inflow, you might need a big data resolution.

Following are scenarios where big data solutions are inherently more suitable [4]

- When you are dealing with enormous semi-structured or unstructured data from multiple sources
- You need to analyze all of your data and cannot work with sampling them.
- The process is iterative (e.g., Searches on Google, Face book graph search).

  Big data solutions work on a fundamentally different architecture which is built the following characteristics (illustrative below):
- **Data Distribution and parallel processing:** Big data solutions work on distributed storage and parallel processing. Simply put, files are broken into multiple small blocks and stored in different units (called racks). Then, the processing happens in parallel on these blocks and the results are merged back together. The first part of the operation is typically called by a Distributed File System (DFS) while the second part is called Map Reduce.
- **Tolerance to failure:** By nature of their design, big data solutions have built-in redundancy. This feature enables big data solutions to scale up even on cheap commodity hardware rather than expensive SAN disks.
- **Scalability & Flexibility:** This is the genesis of entire big data solution paradigm. The user

can be easily added or remove racks from the cluster without worrying about the size for which this solution was designed.

- **Cost-effectiveness:** Because of use of commodity hardware, the cost of creating this infrastructure is far lower than buying expensive servers with failure resistant disks (e.g., SAN).

Big Data in Different Platform

The features of big data used to collaborate with other technologies to serve better. Big data are widely use in cloud computing, healthcare, e-commerce and so on. This section gives an overview of big data with another platform.

a. Big Data Vs. Cloud Computing

The integration of cloud computing, Big Data, and economy of goods and digital services have been fostering the discussion of IT-related services, a huge share of our daily purchasing consumption [5]. It is proposed that the big data applications with 5V features and challenges are and will be driving the explosive advancements of relevant cloud computing technologies in different directions. Figure 3.1 gives a flow of using big data in cloud computing.

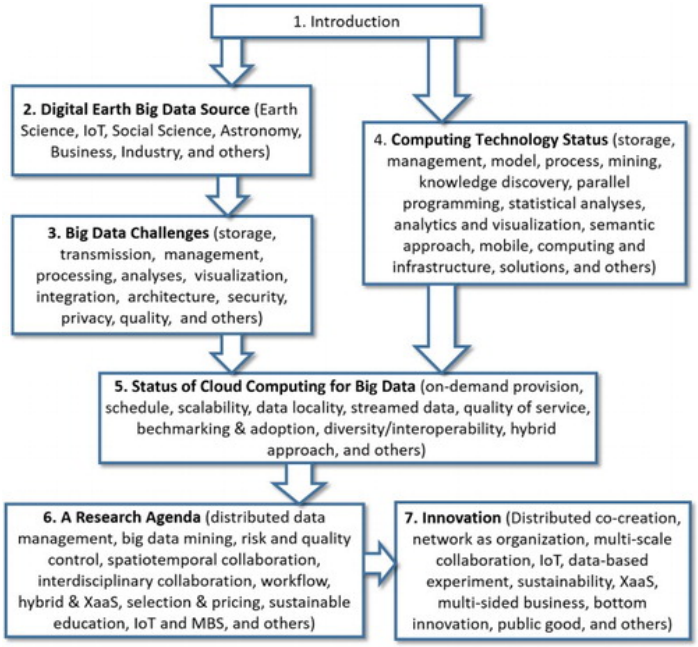

Figure 3.1 Big data and Cloud Computing

b. Big Data Vs. Healthcare

Big data analytics has helped healthcare improve by providing personalized medicine and prescriptive analytics, clinical risk intervention and predictive analytics, waste and care variability reduction, the automated external and internal reporting of patient data, standardized medical terms and patient registries and fragmented point solutions [6]. Some areas of improvement are more aspiration than implemented. The level of data generated within healthcare systems is not trivial. With the added adoption of m Health, e-Health and wearable technologies the volume of data will continue to increase. This includes electronic health record data, imaging data, patient-generated data, sensor data, and other forms of difficult to process data. There is now an even greater need

for such environments to pay greater then attention to data and information quality. "Big data very often means `dirty data' and the fraction of data inaccuracies increases with data volume growth." Human inspection at the big data scale is impossible, and there is a desperate need in health service for intelligent tools for accuracy and believability control and handling of information missed. While wide information in healthcare is now electronic, it fits under the big data umbrella as most are unstructured and difficult to use.

c. Big Data Vs. Other Domain

Big data and the IoT work in conjunction [7] [8]. Data extracted from IoT devices provides a mapping of device interconnectivity. Such mappings have been used by the media industries, companies and governments to more accurately target their audience and increase media efficiency. IoT is also increasingly adopted as a means of gathering sensory data, and this sensory data have been used in medical and manufacturing contexts.

The technology uses Big Data

- EBay.com uses two data warehouses at 7.5 petabytes and 40PB as well as a 40PB Hadoop cluster for search, consumer recommendations, and merchandising.
- Amazon.com handles millions of back-end operations every day, as well as queries from more than half a million third-party sellers. The core technology that keeps Amazon running is Linux-based, and as of 2005, they had the world's three major Linux databases, with capacities of 7.8 TB, 18.5 TB, and 24.7 TB.
- Facebook handles 50 billion photos from its user base.
- Google was handling roughly 100 billion searches per month as of August 2012.
- Oracle No SQL Database has been tested to past the 1M ops/sec mark with eight shards and proceeded to hit 1.2M ops/sec with ten shards.

Conclusion

This paper gives an overview of big data used in different platforms. The features and the characteristics of big data is elaborated in this paper. Big data is an emerging technology which is essential in the current scenario to manage a large amount of data. It is not only used to maintain web-based data but also used to maintain the organizational data, healthcare report of the patients and so on. However, it is main to tighten the security of big data where it stores large data. In case of loss of the storage the alternate solutions have to be proposed and should ensure the confidentiality of the shared data.

References

Agrawal, D, Das, S., & El Abbadi, A. (2011). "Big data and cloud computing: current state and future opportunities". In Proceedings of the 14th International Conference on Extending Database Technologies(pp. 530-533). ACM.

Chen, M, Mao, S., & Liu, Y. (2014). "Big data: A survey". Mobile Networks and Applications, 19(2), 171-209.

Jagadish, H.V, Gehrke, J, Labrinidis, A, Papakonstantinou, Y, Patel, J. M, Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. Communications of the ACM, 57(7), 86-94.

John Walker, S. (2014). "Big data: A revolution that will transform how we live, work, and think".

Jones, S. A. (2017). "Big data and healthcare". Australian Nursing and Midwifery Journal, 25(3), 48.

Katal, A, Wazid, M., & Goudar, R. H. (2013). Big data: issues, challenges, tools and fine practices. In Contemporary Computing (IC3), 2013 Sixth International Conference on (pp. 404-409). IEEE.

Many*ika, J, Chui, M, Brown, B, Bughin, J, D*obbs, R, Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity.

McAfee, A, Brynjolfsson, E., & Davenport, T. H. (2012). "Big data: the management revolution". Harvard business review, 90(10), 60-68.

Web Sources

http://knowledge.wharton.upenn.edu/article/where-analytics-is-headed-next/
http://beckman.cs.wisc.edu/beckman-report2013.pdf
https://www.techrepublic.com/resource-library/whitepapers/big-data-an-analysis-of-large-data/
https://link.springer.com/content/pdf/10.1007/978-3-319-45145-9_2.pdf
https://www.analyticsvidhya.com/blog/2013/07/big-data/
https://www.vidyarthiplus.com/vp/attachment.php?aid=48238
https://en.wikipedia.org/wiki/Big_data_analysis
https://www.slideshare.net/rahimiadel/big-data-sentiment-analysis-awesome
http://www.orafaq.com/aggregator/sources/207
https://rd.springer.com/chapter/10.1007/978-81-322-2494-5_1