

Analytics of Data Science using Big Data

Ch. Sai Krishna Manohar¹

¹(Department of Information Technology, Tirumala Engineering College, Narasaraopet, India)

Abstract : This paper will cover the concepts of Adoption trends of Big Data, needs of BIG DATA, benefits of Big Data and finally the summary, conclusion. My analysis illustrate that the fast-growing and a key enabler for the social business. The paper also discuss about the storing of the data as a Big Data. It deals with the servers that the Big Data is going to maintain at the backend. The insights gained from the user generated online contents and collaboration with customers is critical for customers are critical success in the age of social media.

Keywords – Data warehousing as-a-service (DAAS), Info Sphere Big Insight, Sand Analytic Program.

I. INTRODUCTION

IBM estimates that every day 2.5 quintillion bytes of data are created -- Such that 90% of data in the world today has been created in the last two years. It is a stunning figure and the irony is that we feel less informed in spite of having more information available today. The surprising growth in volumes of data has badly affected today's business. The online users create content like blog posts, tweets, social networking sites and interactions and photos. And the servers continuously log messages about what online users are doing. The online comes from the posts or as an email social media like face book, Gmail, twitter, etc. This data is known as "BIG DATA".

II. What Is Big Data?

Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes are a constantly moving target, as of 2012 ranging from a few dozen terabytes to many PETABYTES of data in a single data set. With this difficulty, new platforms of "big data" tools are being developed to handle various aspects of large quantities of data.

BIG DATA concept means a datasets which continues to grow so much it difficult to manage it using existing database management concept and tools. The difficulty can be related to retrieve the capturing of data, storage, searching and virtualization, etc.

The BIG DATA spans the data across the three dimensions:

Volume: The size of data is very large. It should contain TERABYTES and PETABYTES.

Velocity: It should be used when streaming in to the enterprise in order to maximize its value to the business. The role time is very critical here.

Variety: It extends beyond the structured data, including unstructured data of all varieties.

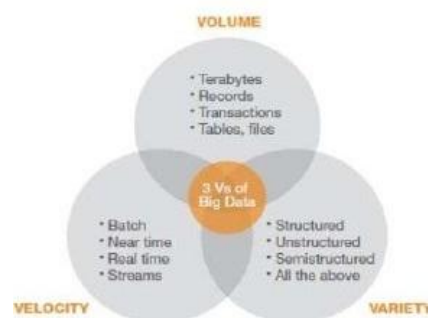


Fig – 1: Velocity, Variety, Volume VS BIG DATA

III. Why Big Data?

When an enterprise can leverage all the information available with the larger data rather than just a subset of its data, then it has powerful advantage over the market competitors. BIG DATA can gain and help to insights to make better decisions.

BIG DATA presents an opportunity to create unprecedented business advantage and better service delivery. It also requires the new way of thinking and new infrastructure about the way of business and IT industry works. The concept of BIG DATA is the way we do the things today.

The International Data Corporation (IDC) predicts that overall data will grow by 50 times by 2020, driven in large part by more embedded systems, such as sensors in clothing medical devices and structures like buildings and bridges. The study also determined that unstructured information – such as files, emails and video – will account for all 90% of data created over the next decade. But the number of IT professionals available to manage all the data will only grow by 1.5 times today's level.

The digital universe is 1.8 trillion gigabytes in size and stored in 500 quadrillion files and its size more than double in every two years time frame. If we compare digital universe with physical universe then it's nearly as many bits of information in the digital universe as stars in our physical universe [7].

IV. Characteristics Of Big Data

A BIG DATA platform can give a solution which is designed specifically with the needs of the enterprise. The following are the basic features of the BIG DATA:

- Comprehensive – It should offer a broad platform, and address all three dimensions velocity, volume and variety.
- Enterprise Ready – It should include the performance, reliability, performance and security features.
- Integrated – It should enable integration with information supply chain including databases, data warehouses and business intelligence applications.
- Open Source Based -- It should be open source technology with enterprise class functionality.
- Low latency
- Robust and reliability.
- Scalability
- Extensibility.
- Allows adhoc queries.
- Minimal Maintenance.

V. BIG DATA OFFERS

There are many vendors offering BIG DATA Analytics are IBM, KOGNITO, etc. Here in this paper I have discussed about the IBM Platform.

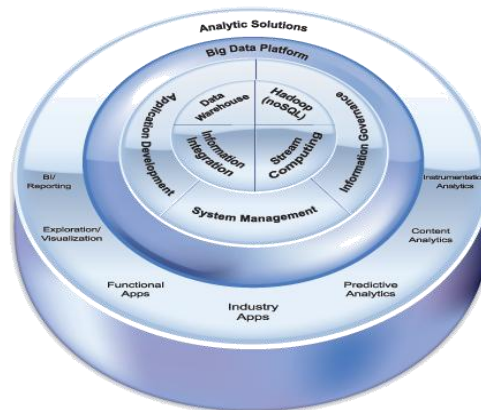


Fig -2: IBM Platform of BIG DATA

VI. ADOPTION TRENDS OF BIG DATA

The analytics of BIG DATA is a fast growing and influential practice. International Data Corporation (IDC) forecast shows that BIG DATA technology and services and market to expected to grow up by \$16.9 billion of data in 2015. It represents a compound annual growth rate (CAGR) of 40% or about 7 times overall the Information and Communications Technology (ICT) market.

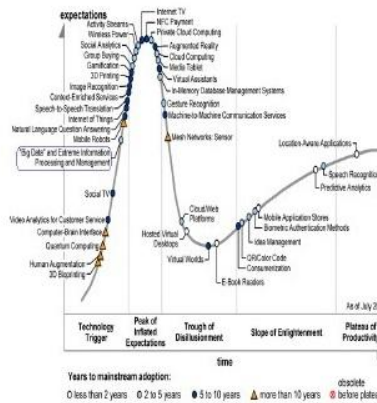


Fig – 3: Gartner Hype Cycle

VII. Benefits Of Big Data

- Data should transfer with high Secure and very fast.
- Stores a vast amount of data.
- The McKinsey Global Institute (MGI) conducted study on BIG DATA in five domain areas, health care in United States, the public sectors in Europe, retail in United States and manufacturing and personal location data –globally. And their study confirms that BIG DATA can generate a data in each of the listed domain areas. For example a BIG DATA can increase its operating margin by more than 60 percent.
- The government administrators in the developed European economy could save more than € 100 billion (8195 rupees) in the operational efficiency improvements using BIG DATA solutions.

VIII. Big Data Challenges

- The main challenges of BIG DATA are data variety, velocity, volume and analytical workload complexity.
- Many organizations are struggling to deal with many problems among the large amount of data. In order to solve this problem, the organizations need to reduce the amount of data being stored and exploit new storage techniques which can improve storage utilization.

IX. Conclusion

Big Data is a new gold rush & key enabler for the social business. A large or medium sized company can neither make sense of all the user generated content online nor can collaborate with customers, suppliers and partners effectively.

Acknowledgements

I hereby declare that I'm very thankful to IEEE, Wikipedia, etc.

REFERENCES

- [1]. <http://www-01.ibm.com/software/data/infosphere/biginsights/>
- [2]. https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&ved=0CD8QFjAC&url=http%3A%2F%2Fgslmug.files.wordpress.com%2F2012%2F02%2Fbig-data-presentation-made-at-the-utah-iseries-user-group-on-feb-82012.ppt&ei=jYI_UaSaJMtrrQe5sYGcG&usq=AFQjCNE5YIz58c49BghBzJ4N8ZiZsd-hzQ&sig2=v4Qg-
- [3]. www.acm.org
- [4]. www.ieee.org
- [5]. W.J. Book, Modelling design and control of flexible manipulator arms: A tutorial review, *Proc. 29th IEEE Conf. on Decision and Control*, San Francisco, CA, 1990, 500-506