



A Big Data Analytics - Challenges With-in New Data, Metadata Management & Analysis Platforms

Mr. Ashish N. Patil¹, Dr. Devendrasingh M. Thakore²

Research Scholar, Computer Engineering Department, BVDCOE, Pune, India ¹

Professor and Head in Computer Engineering Department, BVDCOE, Pune, India ²

Abstract: Nowadays, the information technology spreads very fast due to most generated digital data and the exchanged on Internet day-to-day so large volume of data indicates the new age of the big data management. The world's effective capacity to exchange information through telecommunication network and amount of internet traffic; the data growth has increased up to Exabyte annually. So there are big issue to deal with big data and metadata management. In Big data environment; also need to develop analytical platforms which perform corrective analysis with faster response time. So, the paper is focused on the challenges with-in new data, metadata management and analysis Platforms in a Big Data Analytics.

Keywords: BD (Big Data), BDA (Big Data Analytics), EB (Exabyte), IDC (International Data Corporation).

I. INTRODUCTION

In a digital world, where the data from many resources e.g. telecommunication network contains signal or sensor data or information. The huge data growth due to Internet data which are in many forms. The dataset contains its original data and metadata which means the information about dataset. So for any knowledge discovery the data governance plays an important role in the database engineering.

Big data is collection of large and complex datasets which sizes beyond the ability of commonly used software tools so difficult to analyse within a tolerable elapsed time or small response time; so big data and data analytics are the emerging trends in the information technology domain.

The properties of big data specify with the 5V's model as follows:

- Volume: Data at rest i.e. large volume of data
- Velocity: Data in motion i.e. streaming or real time data
- Variety: Data in many forms i.e. structured, unstructured and semi-structured data
- Value: Data importance i.e. Meaningful data and Validity of data
- Veracity: Data in doubt i.e. finding inconsistencies in data

So, big data has increased the demand of not only the data and information management specialists but also the data analysts. Due to 5Vs (Volume, Velocity, Variety, Value and Veracity) properties of the big data; the Big Data is the big issue in new digital age.

To solve these issue; there is need to develop some suitable and optimized platforms for management of the new data, metadata and information analysis which improve the process of the Big Data Analytics.



Fig 1: 5V's Model to specify properties of Big Data

II. BACKGROUND & MOTIVATION

Within the last two decades, the WWW, more commonly known as the web, has become the main platform for deploying business, social application and organizational information systems. Many organizations have extended the scope of their web-based systems.

As per the report of IDC; developed economies increasingly use data-intensive technologies. 4.6 billion mobile-phone subscriptions worldwide, and between 1 billion and 2 billion people accessing the internet so it marketing of big data is about \$16.1 billion in 2014 it will grow up to \$32.4 billion by 2017 these forecasts indicates that the scope of big data will be grown rapidly in forthcoming future.



Between 1990 and 2005, more than 1 billion people worldwide entered the middle class, which means more people become more literate, which in turn leads to information growth. The world's effective capacity to exchange information through telecommunication networks was 281 Petabytes in 1986, 471 Petabytes in 1993, 2.2 Exabyte's in 2000, and 65 Exabyte in 2007 and predictions put the amount of internet traffic at 667 Exabyte's annually by 2014. Fig. 2 shows an example of World wide data growth recommend by The Big Data Group.

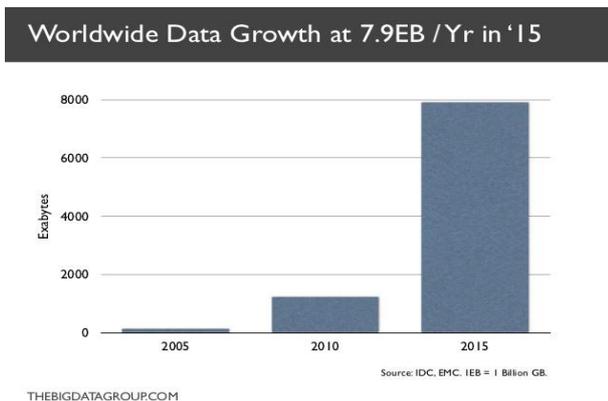


Figure 2: Worldwide data growth

III. RELATED WORK

A. Background of big data and data analytics

Data science introduces the big data and its management. There are some software tools to operate on big data e.g. case study "Apache Hadoop" which is Apache open source software framework for reliable, scalable, distributed computing of massive amount of data; developed in java. Introduction to data mining algorithms which consist of clustering classification and frequent patterns mining technologies. Find solutions to welcome the new age of Big data [1][6].

B. Challenges in big data management

Due to 3Vs properties of big data there are many difficulties in the data management platforms. So there are some techniques to deal with big data management. It also discuss about relationship between data and metadata management. As volume, variety and velocity impacts the complexity of data infrastructures, scaling test environments becomes a significant problem. Requirement for advanced software to realize the potential benefits [4][5][7].

C. Challenges in big data analytics

Review on data analytics studies from traditional data analysis to the recent big data analysis. Multiple type of heterogeneity and poor support for maintaining global consistency of data stores. Focused on performance-oriented and result oriented issues in big data analytics framework and platform [2][16].

IV. ISSUES & CHALLENGES WITH-IN DATA, META-DATA MANAGEMENT & ANALYSIS PLATFORMS

A. Data and Metadata governance

Data in big data environment is with different types of format e.g. structured, unstructured and semi-structured. Metadata can be simply defined as 'information about data within any data environment' which is used to describe the properties of data such as provenance, quality and technical details etc. Metadata management is an important capability that enables managing the change in data while delivering trusted, secure data in the complex big data environment. Metadata management allows extraction of metadata and provides access to metadata. The task of quality management is to assign values to quality attributes. These values are based on the properties of data sets and associated metadata. So there is need to develop good policies for metadata governance.

B. Business Intelligence (BI)

Traditional business intelligence tools typically work with structured data, normally an organization's internal data. BI tools may include business performance management, online analytical processing, statistical analysis and other capabilities for good decision support. It includes Data Mining and Machine Learning Techniques. As Big Data grows to encompass unstructured and semi-structured data, agencies can apply advanced analytics techniques, such as predictive analytics, data mining and natural language processing, to large data sets. So now it is focusing on massively parallel processing.

C. Data Analytics

Analytics means "the scientific process of transforming data into insight for making better decisions". Big Data Analytics (BDA) has a very specific meaning and it refers to doing the analytics work directly in the big data environment. BDA need to analyse the data in real time and often have software applications to run automated predictive models.

V. BENEFITS OF BDA

BDA is used to deal with big data where the statistical and predictive analysis is done by some BI.

- Improve Data integration and quality; easier to integrate and manage complexity
- Improve Business Intelligence i.e. Make better decisions more quickly
- Improve Web analytics
- Reduce security threats and Fraud; more responsive and faster to deliver
- Credit Risk Management
- Enhance transparency and service regarding data governance



- Improve Information Life-cycle Management and productivity

VI. CONCLUSION

In today world practice, due to huge usage of Internet, the digital data is increasing day by day. As increasing the volume of digital data in the world to manage and govern data, there are many challenges in the big data environment. So database administrators are focusing on the advancements of database applications using various techniques. Big data has increased the demand of not only the data and information management specialists but also the data analysts. Due to 5Vs (Volume, Velocity, Variety, Value, Veracity) properties of the big data; the Big Data is the big issue in new digital age. To solve these issue; there is need to build some suitable and optimized platforms for management of the new Data, Metadata and Information analysis which improve the process of the Big Data Analytics.

VII. FUTURE SCOPE

Other important future directions for big data research include the Big Data Science and Technology, BD infrastructure, BD Search, Mining and Management, BD Security & Privacy, BD Applications, BD governance, Information life-cycle management. Finally, future work is needed to improve the model for big data analysis provides smart Business Intelligence for fast decision support and optimizes the performance of Big data analytics, especially since BDA are viewed as the most promising line of approach for achieving universal access in the near future.

ACKNOWLEDGMENT

It is our privilege to acknowledge with deep sense of gratitude to my research guide **Dr. Devendrasingh Thakore**, Professor and Head in Computer Engineering department whose supervision, inspiration and valuable discussion has helped me tremendously in carrying my research work and also thanks to Principal and R&D department for motivating to continue in my research.

REFERENCES

- [1] Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao and Athanasios V. Vasilakos, "Big data analytics: a survey", Tsai et al. Journal of Big Data, A Springer open journal 2015.
- [2] Torsten Priebe, Stefan Markus, Simplity Vienna, Austria, "Business Information Modeling: A Methodology for Data-Intensive Projects, Data Science and Big Data Governance", IEEE International Conference on Big Data (Big Data) 2015.
- [3] J. Alberto Espinosa, Frank Amour, Kogod School of Business, American University, "The Big Data Analytics Gold Rush: A Research Framework for Coordination and Governance", 49th Hawaii International Conference on System Sciences 2016.
- [4] Anuja Kulkarni, Computer Engineering Department, Bharati Vidyapeeth Demeed University-Pune "A Study on Metadata Management and Quality Evaluation in Big Data Management",

- International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653, Volume 4 Issue VII, July 2016.
- [5] Chandra Sukanya Nandyala and Haeng-Kon Kim, School of Information Technology, Catholic University of Daegu, Korea "Big and Meta Data Management for U-Agriculture Mobile Services" International Journal of Software Engineering and Its Applications, Volume 10, No. 2, 2016.
- [6] M. B. Chandak, "Role of big data in classification and novel class detection in data streams", Chandak J Big Data 2016, DOI 10.1186/s40537-016-0040-9, A Springer open journal 2016.
- [7] David Corrigan, IBM InfoSphere, "Integrating and governing big data", IBM Software White paper, January 2013.
- [8] Dan Vesset, Carl W. Olofson, "Big Data in the Enterprise: When Worlds Collide", IDC White paper, February 2014, Sponsored by Oracle and Intel.
- [9] "Improving Media & Entertainment Performance with Big Data", Oracle Enterprise Architecture White Paper, February 2015.
- [10] Megha Kumar, "Engaging in Big Data Transformation in the GCC", IDC White paper, December 2015, Sponsored by IBM.
- [11] Building Governance into Big Data, "A metadata-based approach for ensuring visibility and control for your Apache Hadoop data architecture", A Hortonworks White Paper, February 2016.
- [12] "An Enterprise Architect's Guide to Big Data: Reference Architecture Overview", Oracle Enterprise Architecture White Paper, March 2016.
- [13] Philip Carnelley, Helena Schwenk, "Big Data: Turning Promise into Reality", IDC White paper, October 2016, Sponsored by Dell EMC.
- [14] Rajagopalan M.R, Solaimurugan vellaipandiyan, NRC-FOSS, Centre for Development of Advanced Computing Chennai, India "Big Data Framework for National e-Governance Plan", Eleventh International Conference on ICT and Knowledge Engineering 2013.
- [15] Silberschatz A., Korth H, Sudarshan S., "Database System Concepts", 4th Edition, McGraw Hill Publishers.
- [16] J. Han and M. Kamber, "Data Mining- Concepts and Techniques", 2nd Edition, Morgan Kaufmann, 2006.

BIOGRAPHIES



Engineering.

Dr. Devendrasingh Thakore, Professor and Head in Computer Engineering department at Bharati Vidyapeeth Deemed University, Pune. He has 23 years of experience in teaching field. His areas of interest are Database System and Software



Professor at AGTI's DACOE, Karad. His areas of interest are Distributed System and Big Data Analytics.

Prof. Ashish N. Patil has master degree in Computer Engineering and Research Scholar from BVDUCOE, Bharati Vidyapeeth Deemed University, Pune. He has 6.7 years of experience in teaching field. Currently working as Assistant