# A Study on Big Data Analytics, Approches and Challenges

Shweta S. Lokhande[1], Prof. Rahul Patil[2]

M.E Student, Dept. of Computer Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai,

Maharashtra, India.[1]

Professor, Dept. of Computer Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai, Maharashtra,

India.[2]

**ABSTRACT:** Big data is new paradigm that has been researched in recent years. We have reached in era of interconnectivity among different organization to predict and make decisions in time changing environments. We describe how the big data analytics can prove beneficial for this motive. We see what are steps to approach big data analytics. We also describe what are challenges for applying big data.

**KEYWORDS**: Big data; Big data analytics; Traditional analysis; Big data analysis; Challenges in big data analytics;

## I. INTRODUCTION

An increasing data is being generated every day. The era of such huge amount of data has emerged as a concept called Big data. Big data is heterogeneous and massively generated data. It is voluminous semi-structured or structured data evolving with time. Business routinely collect terabyte of data for analysis. The numbers grows on increasing as corporate allow event logging in more sources, hire more employees, deploy more devices, and run more software. Existing analytical techniques don't do paintings nicely at large scales and typically produce so many false positives that their efficacy is undermined. New termed are coined to handle such data called big data analytics. The big data analytics is far and vast concept .All the drawbacks confronted in traditional analytics are conquered in big data analytics.

From beyond few years there has continued to be no field such as medicine, academics, weather, economy, science that haven't explored big data analytics. Many political events has benefit victory due to analytics of messages from social media websites. The media has reached on pinnacle T.R.P.s due to evaluation of huge facts through BIG Data analytics.. Many studies and researched has boosted velocity due to big data analytics .

The big data analytics along with IoT is used to build smart city with a motive to preserve and revitalize culture[1] .The medical technological know-how cold make use of Big data analytics for pharmaceutical studies to cure illnesses like diabetes, most cancers and so forth[16]. The internet technology used with big data analytics give us self explanatory records rather than easy textual content about surroundings [11].

Due to intercommunication between records, big data analytics is used for Fraud detection. Quick action and accurate decisions in actual time had been feasible because of big facts analytics.

Facebook analytics make use of user data to analyze active users and fan pages. This tool is also used to track interaction between the active users[19].The entrepreneurs make use of Google analytics to track their websites, blogs, viewer etc.[20].

The social analytics in conjunction with big data analytics help us to find context of data being analyzed.

Accordingly Big data analytics is boon on this fast and vast growing industries of data.

This article is organized as follows. Section II defines Big data. In section III we put a light on what is Big data analytics and its purpose. The section IV presents a comparative study between Traditional analytics and Big data analytics. In section V we are going to study the steps to approach Big data analytics. Section VI present challenges in implementing Big data Analytics. Conclusion of this paper is being presented in section VII.

## II. BIG DATA

The Big data is defined as data sets which are so large that traditional technologies unable to process. The amount or size of big data weigh from terabytes, zettabytes , yottabytes and so on. The era of big data started early in 2000s due to introduction of fast and powerful technologies. The one more reason is emergence of storage technologies as the result digital dawn .The data comes in unique formats and nature. The data produced from cell phones, bank transactions, weather senescing devices, log files, text documents, social networking, videos can be structured, semi structured, or unstructured. We need different technologies and methods to process this data. Thus data generated is of different variety. The data generation is growing exponentially with time. The data gathered from social media is best example of data generated with high speed. Hence it is said that big data is of high velocity. The name as big data, gives us an idea that the size and amount is large .The high variety and high velocity is leading in high volume of data. Thus big data is characterized as 3Vs i.e. VOLUME, VELOCITY & VARIETY

The data produced sometimes is unimportant. The sampled data may or may not contain the desired information. Hence we need to aggregate or summarized data to suffice the collected data. We need to perform various data pre processing steps, analyse the data and present it into understandable forms .However when information is in large quantity ,of various variety, and growing with excessive speed, Big data analytics comes into rescue.

## III. BIG DATA ANALYTICS

Big data analytics is the manner of examining large data sets  to reveal undiscovered patterns, unknown correlations, trends persisting in market, customer preferences and other useful business information. The analytical findings can be fruitful into extra powerful marketing, new revenue opportunities, better customer service, stepped forward operational efficiency, competitive advantages over rival organizations and other business benefits.

The big data analytics has been used to answers some questions:
    A.   What happened?
    B.   Why happened?
    C.   What is likely to happen?
    D.   What should be done about it?

To understand big data analytics and answer above questions with help of we take an example of a cosmetics promoting enterprise.

A.   What happened?
After analysing a sales report of cosmetics over a defined period, it appears evidently that there has been decline  in sales  of product. Thus we get clarity of reduction in sales.

B.   Why happened?
The analytics achieved on purchaser critiques, market conditions, it should come with reasons in fall of product income. The solutions could be high charge of merchandise, arrival of new manufacturers, client dissatisfaction and so on.

C.   What is likely to happen?
This stage of analysis is called as prediction stage. Big data analytics are typically used for predicting an occasion. As according to our case we can are expecting that income could attain at top heights in upcoming days or there can be again fall of sales.

D.   What should be done?
The big data analytics carried out over sales can supply us idea to approximately the measures to be taken for enhancing product.

## IV. TRADITIONAL ANALYTICS VS. BIG DATA ANALYTICS

In the traditional analytics data to be processed is known beforehand. The data mostly comes in shape of tables i.e. data is structured. The data from  textual content, Pdfs that are broadly speaking unstructured or semi established can not be managed and subsequently cannot be stored in shape of rows and columns. The RDBMS is a tool used for processing such data. The data follows strict schemas with complicated relationships. The conventional analytical tools and strategies fail to handle big data. The data which is generated varies from gigabytes to yotta bytes is not able to store using traditional data storage systems. The shifting of such data reasons a bottle neck. The single processor could not able to system such huge data. So handling, managing, processing the big data by means of  traditional system is challenging. Consequently traditional analytics has some limitations which have been conquer with the aid of  BIG Data Analytics.

In big data analytics, data to be analyzed may be static or dynamic in nature. The data is constantly being generated over internets and data stored on disks are both considered while analysis. The parallel processing systems are used to process data in big data analytics.. The records saved at extraordinary vicinity is also used for evaluation. The tools like R, HADOOP and  Map Reduce are used for this purpose. . The big data analytical methods can analyze data of any form. The schema used to store data is kept flat. The interrelationship among  them are few and kept simple. The data structure like  Graphs are used to represent data. Therefore big data analytics proves more powerful than traditional data analytics.

## V.  STEPS TO APPROACH  BIG DATA ANALYTICS

The analysis of  big data can be performed as follows:
- A.  Collection
- B.  Organization
- C.  Analysis
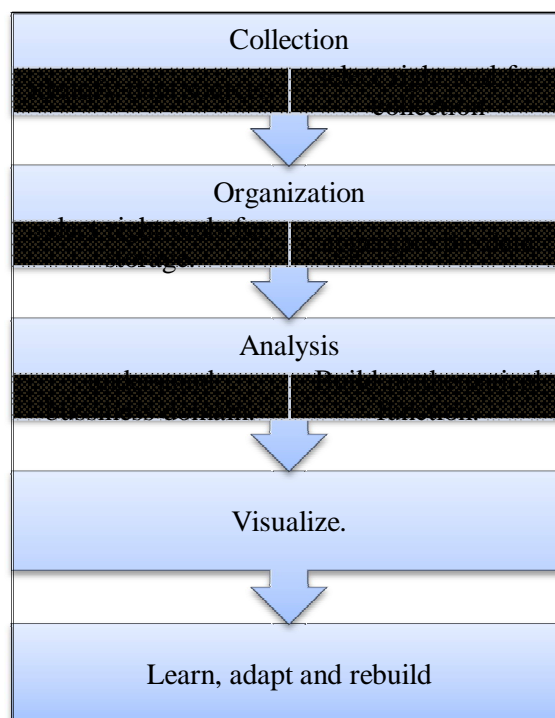- D.  Visualize
- E.  Learn, Adapt, Rebuild



Fig.1. Steps to approach big data analytics.

*A.* *COLLECTION:*

It is basic steps for performing analytics. The data on which analytics is performed must be collect from different sources. Figuring out proper supply can help to particular analysis. One should select right gear and techniques to acquire records.. One can collect data over internets, customer forms, interviews or interrogation, proper feedbacks etc. Selection of right tools can help to save times.

*B.* *ORGANIZATION:*

The data collected need to be organized in proper forms. The data must be cleaned before organization. The data organization or modeling can be  like data ware houses and schematics webs.

*C.* *ANALYSIS:*

- *Understand business domain:*

The knowledge of business for which analytics is done , can help to guide analytics for desired results. The understanding of commercial enterprise for which analytics is computed, can help to manual analytics for desired consequences. One must have a good studies or look at should be carried before analytics. for example to perform affected person fitness analysis, one have to be aware of his preceding fitness records, scientific time period expertise, etc. The domain  knowledge and application knowledge proves beneficial for designing  algorithms and set methods for analytics [12].It becomes unclear to find effective measure, if  we do not have properly identifying correct tools and techniques:

The identification of tools and technique foster the analytics. The tools depending upon the application such as desktop, clouds must effectively selected. The data analyst must cleverly choose the architecture for analysis. For e.g. statics analytics can be carried out with help of R or MATLAB where as dynamic analysis which occurs in changing environment can be carried in HADOOP architecture. The overview of tools[10] provide us with performance of various tools available for Big data analytics which can be help to save computational time and cost.

- *The building of mathematical functions:*

To carry out analysis we need to follow specific algorithms. This algorithms optimize the results .One can yield better performance of system using correct mathematical functions and algorithms.

*D.* *VISUALIZATION:*

As the data set is very large, one can frequently zoom and note each detail. The table layout doesn't show useful at such example.. The visual aids or demo graphs like pie charts,  which are beyond static dash board where users can note and can drag any widgets[5]  are used as visualization tools. Visualization provide a way to maintain context by showing co-related variable .It is relevant to display relation of various streaming data and came help to identify patterns[6].

*E.* *LEARNING , ADAPTING AND REBUILDING:*

The analytics over huge information help us to analyze the causes effects over an occasion. You can come up with decisions and predictions using the result of evaluation.

The new technologies can be used to adapt the  upon the results or decision. Accordingly, the technique of collection, organization , and analytics goes on and on. depending upon the consequences or choice , we want to rebuild the strategies .

## VI. CHALLENGES IN BIG DATA ANALYTICS

We need to cope up with   numerous demanding situations  to realize its true potential, although the application of big data analytics to security problems has significant promise .As per requirements are generated for sharing the data among the industries, privacy is main concern, as sharing data goes against the privacy policy of companies. Advancements in big data analytics, have given us tools to extract and correlate this data, making privacy violations easier. Therefore, development must be done in big data applications with an understanding of privacy principles and recommendations though privacy regulation  exists in some sectors. As the application of big data analytics to security problems has significant promise, many parties, including industry that can use our information for marketing and advertising, government, and  sophisticated criminals like intruder or attacker are attracted to large storage of data. Therefore, the role of big data application architects and designers is to be proactive in creating safeguards to prevent abuse of these big data stores[15].

As   IoT is best source of big data, the hardware cost become burden. More and more data storage technologies need to build to store large data sets as data generation is a continues process .Along with this,  processor architecture  need to be revised to turned them to handle such big data. The analytical tools are made powerful to produce correct and accurate results. The visualization tools must deliver the results in user understandable forms.

As surveyed [4], the leading barriers to big data analytics   are inadequate staffing and skills. After all, many organizations are still new to big data analytics and their  skill sets are  not quite the same as that for business intelligence and data warehousing, for which most organizations have developed their skills. Problems with database software can form a bottle neck to big data analytics. Issues arise when the current database software lacks in-database analytics , has scalability problems with big data that are they can't process analytic queries fast enough , or cannot load data fast enough .

## VII.    CONCLUSION

The goal of this paper is to describe, and mirror on big data analytics. The paper firstly defines what is meant by big data. We presented various definitions of big data, highlighting the fact that size is only one dimension of big data along with this, other dimensions such as velocity and variety are equally important. The paper's primary cognizance has been on analytics to advantage legitimate and valuable insights from large information. We additionally researched the difference between conventional analytics and large records analytics. The steps to approach analytics has been explained in this paper. The challenges are addressed which might be faced even as imposing big data analytics.

## REFERENCES

1.  Yunchuan Sun, Houbing Song, Antonio J. Jara," Internet of Things and Big Data Analytics for Smart and Connected Communities", IEEE Access, Vol. 14, No. 8, August 2015.
2.  Surend Raj Dharmapal, Dr. K.Thirunadana Sikamani,"Big Data Analytics Using Agile Model", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016, 978-1-4673-9939-5/16/$31.00 ©2016 IEEE.
3.  Hu Shuijing,"Big Data Analytics-Key Technologies ans Challenges", 2016 International Conference on Robots & Intelligent System, 978-1-5090-4155-8/16 $31.00 © 2016 IEEE,DOI 10.1109/ICRIS.2016.30.
4.  Philip Russom," Big Data Analyitc", Fourth Quarter, 2011.
5.   Datameer,"The Guide to Big Data Analytics",2013.
6.  Danyel Fisher,Rob DeLine,Mary Czerwinski,Steven Drucker," Interactions with Big Data analyitics.",
7.  Amir Gandomi,Murtaza HaiderTed," Beyond the hype: Big data concepts, methods, and analytics", International Journal of Information Management 35 (2015) 137–144.
8.  Ms. Vibhavari Chavan, Prof. Rajesh. N. Phursule," Survey Paper On Big Data", Vibhavari Chavan et al,  (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7932-7939www.ijcsit.com.
9.  Vijaylakshmi S. and Priyadarshini J.," Big Data Analysis Based On Mathematical Model: A Comprehensive Survey", Vol. 10, No. 5, March 2015 Issn 1819-6608,ARPN Journal of Engineering and Applied Sciences.
10.  Hadi Hashem and Daniel Ranc," A Review of Modeling Toolbox for Big Data", International Conference on Military Communications and Information Systems ICMCIS (former MCC)Brussels, Belgium, 23rd - 24th May 2016.
11.  Claudia Vitolo , Yehia Elkhatib , Dominik Reusser , Christopher J.A. Macleod ,Wouter Buytaert ," Web technologies for environmental Big Data", Environmental Modelling & Software 63 (2015) .
12.  Xindong Wu1, Xingquan Zhu , Gong-Qing Wu , Wei Ding," Data Mining  with Big Data", Proc.in IEEE.
13.  Shuang Wang, Luca Bonomi, Wenrui Dai, Feng Chen, Cynthia Cheung, Cinnamon S. Bloss, Samuel Cheng," Big Data Privacy in Biomedical Research", Pro. In IEEE.
14.  Paul C.Zikopoulos, Chris Eton, Thomas Deutch, "Understanding Big data- Analytics For Enterprice Class,Hadoop and streaming Data."

15. Alvaro A. Cárdenas ,Pratyusa K. Manadhata,Sreeranga P. Rajan, "Big Data Analytics for Security", 1540-7993/13/$31.00 © 2013 IEEE.
16. Jamie Cattell, Sastry Chilukuri, and Michael Levy," How big data can revolutionize pharmaceutical R&D", April 2013.
17. http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics.
18. https://en.wikipedia.org/wiki/Big data.
19. https://blog.kissmetrics.com/guide-to-facebook-insights/
20. https://moz.com/blog/absolute-beginners-guide-to-google-analytics.