A Comparative Study on Big Data and Big Data Analytics (BDA) and Challenges

Smt. Nitnaware Reshma Ramdas[1,] Dr. Ajit A. Masleker[2]

[1]Research Scholar, Department of Computer Science & IT, Dr. B.A.M.U, Aurangabad-431004(MS), India

[2]K. T. P. Computer of Science College, Osmanabad Maharashtra – 413606(MS), India

Abstract-This paper gives a clear idea about the is big data And Big Data Analytics. Big data refers not only to large data sets, but also to high speed and variety that are difficult to process with traditional tools and techniques. Such data is growing rapidly, so we need to find some way to extract key insights and value from these datasets. Decision makers also need to derive valuable vision from large, constantly changing data, from day-to-day transactions to customer interactions to data in their social networks. Such a vision can be communicated using big data analytics, the application of advanced analytical techniques to big data. This white paper aims to explore some of the different analytical methods and tools that can be applied to big data, and the costs incurred in applying big data analytics in various decision-making domains.

Keyword: - Big data Analytics (BDA), Hadoop , Velocity , Variety, Analytics, decision-making.

INTRODUCTION TO BIG DATA

Big Data is Simple words because is collection of large volumes of data with Different Velocity and Different Variety of data asset. Velocity is speed and variety means different formats (like audio, video, image etc.). A big Data it is key for analyze systematically extract information from, or otherwise deal with data records (sets) that are too large or complex to be distributed with by traditional tools data processing software. Big Data is a database stored unlimited data into form of records.

Big data analytics allows data scientists and various other users to evaluate large volumes of transaction data and other data sources that traditional business systems would be unable to tackle. Traditional systems may fall short because they're unable to analyze as many data sources. Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.

Lots of data is being collected and warehoused, Web data, e-commerce, purchases at department/grocery stores, Bank/Credit Card transactions, Social Networks. Big data refers to data sets whose size is beyond the ability of typical database software tools to "capture, store, manage and analyze."

BIG DATA ANALYTICS

Today, people don't just want to collect data, they want to understand its meaning and importance, and use it to take it into account when making decisions. Data analytics is the process of applying algorithms to analyze data sets to extract useful and unknown patterns, relationships, and information. Use data analytics to extract previously unusable and valid hidden patterns and information from large data sets. Big data analytics is of great importance in an era of data overload and can provide unexpected insights and benefits in overall decision-making in various sectors. Big Data analytics are competent to provide scientific, technical, and humanitarian levels of competence.

BIG DATA ANALYTICS TYPES

There are Four Types of Big Data Analytics
1. Descriptive Analytics :
This type of Big Data Analytics that provide data as a descriptive manner means the big data as the past data provide to easily read peoples.
2. Diagnostic Analytics :
This type of Big Data Analytics why to the complication occur & then how to Handling, BDA use the different techniques like Data Mining, Data Recovery etc.
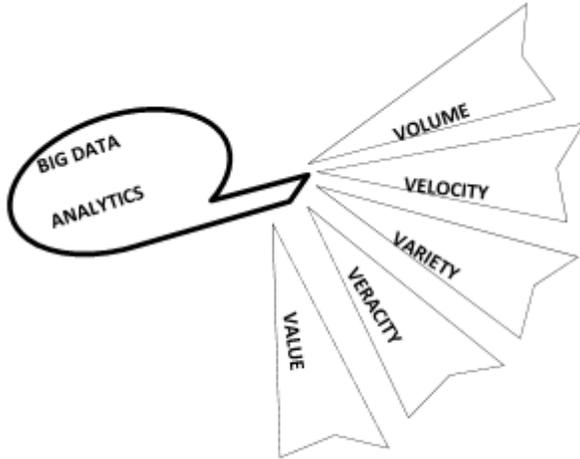
Fig : BDA Characteristics

3. Predictive Analytics :

This type of Big Data Analytic Use future outcome using past data (Existing Data). It is analytic technique for designing Models to the future outcome.

4. Prescriptive Analytics :

This type of Big Data Analytic use to build an algorithms for prescriptive, these analytics might use computer vision, ML (machine learning), NLP (natural language processing), signal processing, image processing, speech recognition and more techniques to advocate analysis.

BIG DATA ANALYTIC CHARACTERISTIC

- Volume : The First V of Big data is Volume it is just like base of the big data. This is used to initialize size and amount of data that is collected. If the volume of data is very large then it is actually be considered as a Big Data or not is depend upon volume of data.
- Velocity : The next V is Velocity that refers as High SPEED of accumulation of data.
- Variety : The next V is Variety that refers as different formats data they are generated or collected in different ways like as Facebook, twitter, telegram, YouTube etc.
- Veracity: The next V is Veracity that refers data accuracy and true worthiness means quality of data (consistency).
- Value: The Last but not Least V is Value that refers Value is the worth of the data being collected. Means the bulk of data having no good to any company, unless you turn it into something

useful. And the value of data is like detect disease after better treatment and reduce cost.
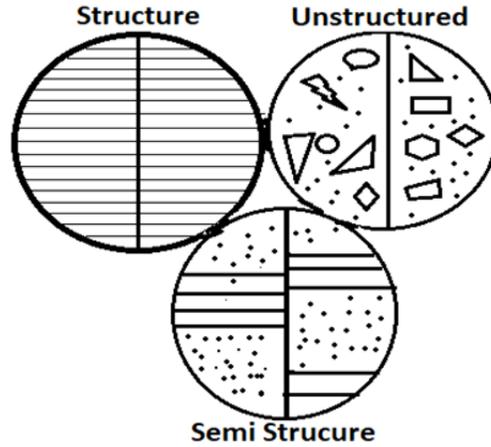
BIG DATA ANALYTICS TYPES

Fig : Types of Big Data Analytics

- Structure: -
The Structure Big Data type the data are return to the record format means data has defined the length and format of data (row & column). These type of data stored in the database and it is easily accessed by Human & Programs.
- Unstructured: -
The Unstructured Big Data types means Data refers to unorganized data they not contains rows and columns, they contains audio, video, images, pictures etc. not relative to the database.
- Semi-Structure:-
The Semi-Structure Big Data which was built specifically for DBMS, may store semi-structured data. XML may render data that is semi-structured. Users can modify the attributes, tags, and elements in XML, which also helps to store the data in a hierarchical format.

BIG DATA AND BIG DATA ANALYTICS

Big data is not limited to complex, massive or huge data, it is identified by its characteristics. The ubiquity of complex data generated from new data sources has evolved into a new technology called Big Data Analytics. It is about analytical workloads that are associated with some combination of data volume, data velocity, and data variety that may include

complex analytics and complex data types. Big Data Analytics has shown its potential in many application domains. BDA has also achieved success to fulfill the modern users' demands up to some extent. Big Data analytics is the process of examining large and complex data sets that often exceed the computational capabilities.

BIG DATA ANALYTICS CHALLENGE

- BIG:- The challenge you are attempting to resolve turns fundamentally as your set of data is big, large. What algorithm is the most effective for the assignment? How can we collect reliable RAM to processing the set of data? What is the most effective sampling strategy to assure a representative subset?
- DATA :- The difficulty is that most data need cleaning. Cleaning is most often increased data set.
- ANALYTICS :- Analytics is a branch of computer science that looks for useful patterns in data using arithmetic, statistics, and machine learning. Analytics, often known as data analytics, is the process of sorting through enormous data sets to find, understand, and communicate new information. Not creating an accurate model or acquiring good, clean data is the most challenging aspect of data science.
- Need for Understanding Data:- Without understanding data we cannot analysis we need first understand, means we should have good knowledge of the type data we are dealing with , sources of data sets from which they are coming in and what should be derived from the data as the result.
- Domain Knowledge: We need good knowledge on the domain of which target data is related to understanding the customer's behavior is a key writing particular domain.
- Accuracy of Data : This is related to data quality means data comes from different sources, the are at most chances for junk in them as well. We have need to data processing and analyzing of less junk.
- Data Integration: We must need to analyzing a connect well analyzed, meaningful, well defined data , means to the quality of meaningful data.

- Data Volume: Analyzing and producing insights to the customer in a timely manner from huge volumes of data is the real challenge.
- Methods / Tools used: Big data analytics provide more methods and tools Regression analysis, Monte Carlo simulation, Factor analysis, Cohort analysis, Cluster analysis, Time series analysis, Sentiment analysis and so on.

CONCLUSION

Data is extremely valuable today and growing exponentially. We need to find insights from data for a variety of purposes. Due to the large amount of mixed data, it is impossible for traditional data mining tools to process this data and generate real-time analytics. Therefore, big data is applied to process such large amounts of heterogeneous data in real time and describes the main challenges facing big data analytics. Tools and techniques available for big data such as the Hadoop ecosystem. We also propose available solutions to address these analytical challenges. This Paper will be useful for researchers working in the field of big data and big data analytics. These tools can be used in various fields where data analytics is required. Many more tools have been introduced in the market and the existing products are also under constant improvement. The demand for better analytics tools is increasing constantly which is only going to increase further in future.

REFERENCES

[1] Simon Walkowiak ,"Big Data Analytics with R Book ", July-2016

[2] C. Dobre, F. Xhafa. Intelligent services for Big Data science. Future Generation Computer Systems 37 (2014) 267–281

[3] Eckerson, W. (2011) "BigDataAnalytics: Profiling the Use of Analytical Platforms in User Organizations," TDWI, September. Available at http://tdwi.org/login/default-login.aspx?src=%7bC26074AC-998F-431BBC994C39EA400F4F%7d&qstring=tc%3dassetpg

[4] Lekha R. Nair , Sujala D. Shetty,"Research in Big Data and Analytics: An Overview" International Journal of Computer Applications (0975 – 8887) Volume 108 –No 14, December 2014

[5] Blog post: Thoran Rodrigues in Big Data Analytics, titled "10 emerging technologies for Big Data", December 4, 2012.

[6] Douglas, Laney, "The Importance of 'Big Data': A Definition". Gartner. Retrieved 21 June 2012.

[7] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," interactions, vol. 19, no. 3, pp. 50–59, May 2012

[8] Ari Banerjee senior analyst, heavy reading, "Big data and advanced analytics in Telecom: A Multi-BillionDollar Revenue Opportunity," December 2013.

[9] Nada Elgendy, Ahmed Elraga ,"Big Data Analytics: A Literature Review Paper", August 2014 , DOI: 10.1007/978-3-319-08976-8_16 .

[10] Nikhil Madaan, Umang Kumar, Suman Kr Jha "Big Data Analytics: A Literature Review Paper" June 2020, DOI:10.17577/IJERTCONV8IS 10003.

[11] SimpliLearn , Accessed by 20 Dec 2022 'What is Big Data Analytics and Why It is Important?' <https://www.simplilearn.com/what-is-big-data-analytics-article>

[12] Sarah Al-Shiakhli ,"Big Data Analytics: A Literature Review Perspective", Information Security, master's level (120 credits) 2019.

[13] Mrs.Reshma Ramdas Nitnaware," Big Data Analytics Technology and Machine Learning in Big Data ", April-2019 (47-50) (Nitnaware, 2019).

[14] Hemant Kumar Singh,Vinodini Katiyar,"CHALLENGES OF BIG DATA ANALYTICS", May 2021 (162-169).

[15] www.Google.com

[16] www.Wikipedia.com