



Journal Homepage: - www.journalijar.com
**INTERNATIONAL JOURNAL OF
 ADVANCED RESEARCH (IJAR)**

Article DOI: 10.21474/IJAR01/5686
 DOI URL: <http://dx.doi.org/10.21474/IJAR01/5686>



RESEARCH ARTICLE

TECHNIQUES OF BIG DATA ANALYTICS: A LITERATURE SURVEY.

G. Kalpana.

Asst. Professor, Dept of MCA, CBIT.

Manuscript Info

Manuscript History

Received: 21 August 2017
 Final Accepted: 23 September 2017
 Published: October 2017

Key words:-

Big Data, 5 V's, data set, Data Analytics.

Abstract

These days' Big data influences lot of changes in the field of business world. Big Data describes the large volume of data – both structured and unstructured. It is used to handle not only large volume of data but also complex data. This paper explores what exactly Big Data is and how it is used in different applications. This paper also includes brief information about Big Data, five V's of Big Data, importance of Big Data, the list of few popular tools presently available in the market to process Big Data. It also includes the brief information about Big Data Analytics and few techniques to analyze the Big data effectively.

Copy Right, IJAR, 2017,. All rights reserved.

Introduction:-

Big Data is a very familiar term which describes voluminous amount of data that is structural, semi-structural and Sub structural data that has potential to be mined for information. It does not refer to any specific quantity, but this term is often used when speaking about the petabytes and Exabyte of data.[1]

The important properties of Big Data are Volume, Variety, Velocity, Variability and Value [1]. The **volume** of Big Data is increasing day by day. The large amount of data accumulated through sensor networks and social websites rising from petabytes to Zetabytes. **Variety** of Data produced from different categories, consists of structured, semi structured unstructured, semi structured and raw data which are very difficult to handle with traditional systems. **Velocity** is the speed at which the data is generated and become historical. Big data handles the outgoing and incoming data rapidly. **Variability** is the amount of variance used in summaries kept within the data bank and refers how they are closely clustered within the data set. **Value** All e-commerce systems and enterprises try to improve the relationship with customers by providing value added services. The study on customer trends and behaviors in the market are analyzed. Users can also query the data store to find out business trends and they can change their master plan or strategies. Since big data is open to all, it creates functional analysis.[2]

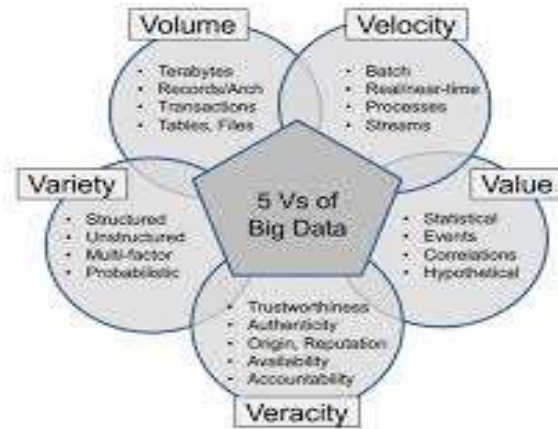


Fig 1:-Five V's of Big data

Importance of Big Data:-

The importance of big is not only how much data you have, but what we can do with it. We can take data from any source and analyze it to find answers which enable 1) cost reductions, 2) time reductions, 3) new product development and optimized offerings, and 4) smart decision making. When we combine big data with high-powered analytics, we can accomplish business-related tasks such as: Finding root causes of failures, issues and defects in near-real time, Generating coupons at the point of sale based on the customer's, buying habits, Recalculating entire risk portfolios in minutes, Detecting fraudulent behavior before it affects your organization.[3]

The processing of Big Data:-

The steps for processing of big data include *Storage ,Formatting, Cleaning ,Data Understanding ,Data Integration ,Data Access ,Data Analysis ,Data Visualization.*

The simple and actual data is stored in database suitable for respective application. We need to extract the knowledge from given data, for that first we make some formatting changes in actual data and then clean it for unnecessary storage anomalies. After doing this we can understand what actual data is all about and depending on that we can access it and can integrate it wherever and whenever required. Data accessing and analysis together will give an idea about how data can be visualized. [6]

Techniques For Analysing Big Data:-

Association rule mining:-

which is a used to discover correlations between variables in large databases. It was first used in point of sales system to discover interesting relations between products. Association rule mining is used to organize the products in better proximity to each other which gives benefits to organizations in getting increase of sales, extract information about various visitors to websites from web server logs and we can also analyze the biological data to find out new relationships and monitor system logs to detect intruders and their malicious activity.

Genetic algorithms:-

Are inspired through the mechanisms such as inheritance, mutation and natural selection. Genetic algorithms are used to schedule doctors for hospital emergency rooms. They are also used to develop fuel-efficient cars and also to generate "artificially creative" content such as puns and jokes.

Statistical classification:-

Is a method to identify new observation categories. Based on the historical data(training set) it identifies the new observation category. Statistical classification is used to automatically assign documents to categories and categorize organisms into groupings. It is also used to develop profiles of students who take online courses.

Machine learning:-

Is used to devise complex models and algorithms that helps in predictive analysis which helps researchers, data scientists, engineers, and analysts to "produce reliable, decisions and results" to uncover "hidden insights". It gives

computers the ability to learn without being explicitly programmed. Machine learning is used to help distinguish between spam and non-spam email messages, learn user preferences and make recommendations based on this information. Machine learning also determines the probability of winning a case, and setting legal billing rates.

problems in using machine learning methods for big data analytics is similar to traditional data mining algorithms which are designed for sequential or centralized computing. One of the most possible solutions is to make them work for parallel computing. [10]

Regression analysis:-

Involves manipulation of some independent variables from the influences of dependent variable. It describes how the dependent variable changes when the independent variable is varied. It is effective with continuous quantitative data like weight, speed or age. Regression analysis is used to determine how the customer satisfaction affect customer loyalty •For example " the number of supports calls received may be influenced by the weather forecast given the previous day" .

Sentiment analysis:-

Helps researchers to determine the sentiments of speakers or writers with respect to a topic. Sentiment analysis is used to help to improve service at chain of hotels by analyzing feedback of guests. So that they can decide incentives and services for customers. IBM has developed IBM Social Media Analytics [9] is a powerful SaaS solution. It captures both structured and unstructured data from social networking sites and also helpful in studying the attitudes, opinions and trends. Later it applies tools of predictive analysis to find customer behavior and improve customer experience. This helps the company to create personalized campaigns and promote to increase the consumer base.[4]

Social network analysis:-

Techniques were first used in the telecommunications industry, and later on used by sociologists to study interpersonal relationships. They are now being applied to analyze the relationships among people in different fields and commercial activities. Social network analysis is used to see how people from different places interact with each other. It is also used to determine fraudulent activities.[4] social network is also useful to know the likes and dislikes of different people.[8]



Fig2:- Advantages of Big data[5]

Big Data Analytics:-

Big Data Analytics, is the process of examining large data sets that containing different data types i.e., big data is to uncover all hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. Then analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over rival organizations and other business benefits.

Big data analytics helps companies make more informative business decisions and enables data scientists, predictive modellers and other analytics professionals to analyze large volumes of transactional data, as well as other forms of data that could be untapped by conventional Business Intelligence(BI) programs which also includes web server logs, Internet click stream data, social media content, social network activity reports, text from customer emails, survey responses, mobile phone call detail records and machine data captured by sensors and connected to the Internet of Things.

From the beginning, Google, LinkedIn, and eBay and Face book were built around big data. These firms did not have to integrate big data with more traditional sources of data and the analytics performed upon them since they did not have to merge big data technologies with their traditional infrastructures. [1]

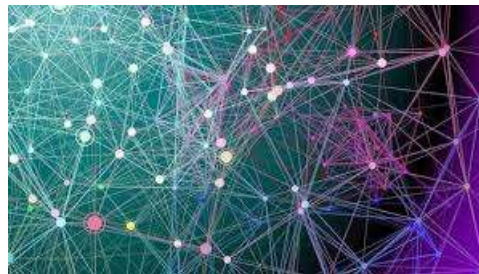


Fig3:- Big data challenges[5]

Big Data Challenges:-

1. The challenge is how to deal with the size of Big Data.
2. It means combining multiple Data Sets. The challenge is how to handle various sources, types, and formats.
3. One of the important challenges is how to react to the flood of information in the time required by the application.
4. How can we cope up with uncertainty, imprecision, missing values and untruths?
5. How good is the data? How broad is the coverage?
6. How fine is the sampling resolution? How periodically are the readings?
7. How well understood are the sampling biases?
8. Is there data available, at all?
9. This is a huge challenge to find out high-quality data from the tons of data.
10. The challenge is determining the quality of data sets and relevance to particular issues.
11. Are there areas without coverage? What are the implications?
12. Th management challenges like Data privacy, Governance, Ethical, Security.[2]

Tools:-

1. *NoSQL Databases* MongoDB, CouchDB, Cassandra, Redis, BigTable, Hbase, Hypertable, Voldemort, Riak, ZooKeeper.
2. *MapReduce* Hadoop, Hive, Pig, Cascading, Cascalog, mrjob, Caffeine, S4, MapR, Acunu, Flume, Kafka, Azkaban, Oozie, Greenplum
3. *Storage* S3, Hadoop Distributed File System
4. *Servers* EC2, Google App Engine, Elastic, Beanstalk, Heroku
5. *Processing* R, Yahoo! Pipes, Mechanical Turk, Solr/Lucene, ElasticSearch, Datameer, BigSheets, Tinkerpop



Fig4:- Big data applications In realtime[5]

Applications of Big Data:-

Big Data tools and Technologies are helpful for companies to interpret the huge amount of data very faster which also helps to production efficiency and also to develop new data-driven products and services. So, Big data applications are creating a new era in industry.

1. Retail/Consumer:
 1. Analytics of supply chain management
 2. Event and behaviour based targeting
 3. Market and consumer segmentation

2. Finance and Fraud Services:
 1. Fraud detection and security analytics
 2. Risk analysis and management
 3. Abnormal trading pattern analysis
3. Health And Life Science:
 1. -Clinical trials data analysis
 2. Disease pattern analysis
 3. -Drug discovery and development analysis
4. Telecommunications:
 1. Telephone call detail analysis
 2. Network performance and optimization
 3. Customer churn prevention
5. Ecommerce and Customer service:
 1. Event analytics
 2. getting right offer at the right time
 3. Cross channel analysis
 4. Next best offer [7]

Conclusion:-

This paper explores the steps to process Big Data, Techniques for analyzing the Big Data. So that we can apply in different fields. Big data is used in many real time applications like banking, finance, health care, marketing, stocks, agriculture, chemistry, data mining, satellite, astronomy, cloud computing, Mobile network service etc. further we can also use these big data analytics in new fields like print media, astrology etc. this paper helps the beginners to get ideas related to their field of research.

References:-

1. Kuchipudi Sravanthi, Tatireddy Subba Reddy on “Applications of Big data in Various Fields” IJCSIT 2015
2. Priya Parhate¹, Gaurav Ghogle², Jyoti Bhange³, Ashwini Ingle⁴ on “review paper on big data: challenges and applications” IRJET January 2017
3. https://www.sas.com/en_us/insights/big-data/what-is-big-data.html
4. Samiddha Mukherjee¹, Ravi Shaw² on “Big Data – Concepts, Applications, Challenges and Future Scope” IJARCCCE Feb. 2016
5. <https://www.google.co.in/search?q=big+data+pics>
6. prof. a. r. wasukar¹, prof. p. a. pawade² “a review on what is big data and how to handle the enormous data through big data “
7. <https://www.greycampus.com/opencampus/big-data-developer/applications-of-big-data>
8. <https://www.firmex.com/thedealroom/7-big-data-techniques-that-create-business-value/>
9. <http://www-01.ibm.com/software/analytics/solutions/customer-analytics/social-media-analytics/>
10. Chun-Wei Tsai¹, Chin-Feng Lai², Han-Chieh Chao^{1,3,4} and Athanasios V. Vasilakos⁵ on “Big data analytics: a survey”.