Big Data Analytics - Tools, Challenges, Applications, Future Scope

Kavyashree E D

Assistant Professor, Department of Computer Science and Engineering ATME College of Engineering Mysuru Mysuru, India

---***---

Abstract:

Growth in various technologies in this digital era has led to increase in data day by day exponentially. Big data analytics is one of the trending techniques to obtain useful information from big data. These data are available in structured, semi structured, and unstructured format. Big data encompasses of 3V 's, Velocity, Volume and Variety. It's difficult to process and manage huge data with traditional tools so advance tool is needed. Big data might be petabytes or Exabyte's consisting of million billion and trillion information of users from various sources eg mobile data, sales data, social media, web and so on. Big data is used for better understanding customer's requirement and preferences. As users are increasing day by day there is no end in increase in data its everlasting process in days to day routine. For any business workflow big data analytics is essential part.

Keywords: Big Data, Big Data Analytics, Tools, Integration.

1. INTRODUCTION

Big data is one of the strongest issues in today's era. In IT's world data is everything and it is getting multiplied by the users every day. Earlier data was in kilobytes and megabytes but now it is in the form of terabytes and even more. Data is of no use until it turns into useful information. Dimensions of big data are volume, variety, and velocity. [1][2][3]
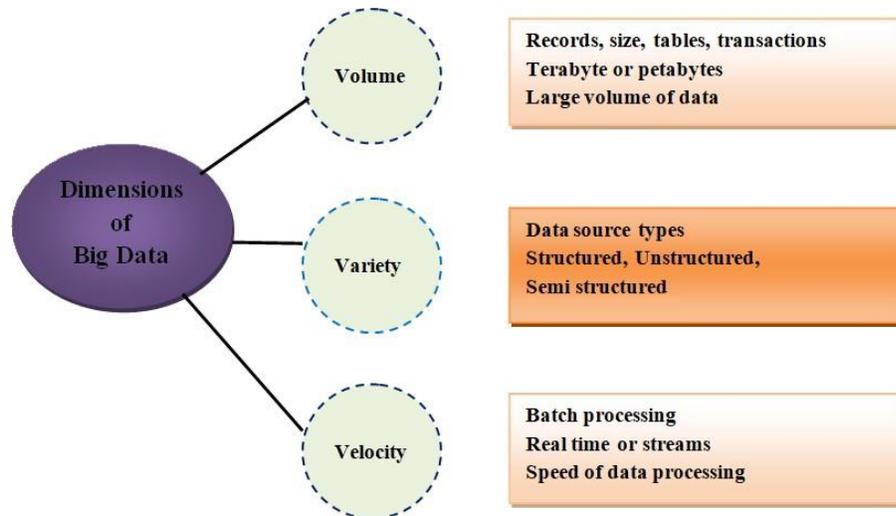

Fig 1.1: Dimensions of big data

- **Volume:** Storage of data requires space; organization has to scale hardware and software in order to occupy new data. It determine the quantity of generated and stored data. Data size determines the value and potential

insight and whether it can be considered for big data or not.

- **Variety:** Type and nature of data such as text, video, audio it also completes missing pieces through data fusion.
- **Velocity:** The speed at which data is generated. Focuses on frequency of generation and frequency of handling. [4][5]

2. TOOLS IN BIG DATA

Key providers of big data services are IBM, Oracle, SAP, Hewlett Packard, Dell Incorporation, Teradata, Splunk, Amazon web services[2].Modern tools are needed for storing, analyzing, reporting and to work more on data some advanced tools are needed.Big data tools are apache Hadoop, Cloudera distribution for hadoop, Cassandra, Knime, DataWrapper, MongoDB, Lumify, HPCC, Storm, Apache SAMOA, Talented, Rapidminer, Qubole, Tableau, R, Elasticsearch, OpenRefine, StataWing, CouchDB, Pentaho, Flink, Data Cleaner, Kaggle, Hive, Spark, IBM SPSS Modeler, OpenText, Oracle Data Mining, Teradata, BigML, Silk, CartoDB, Charito, Plot.ly, BlockSpring, Octoparse, Neo4j[1][6]
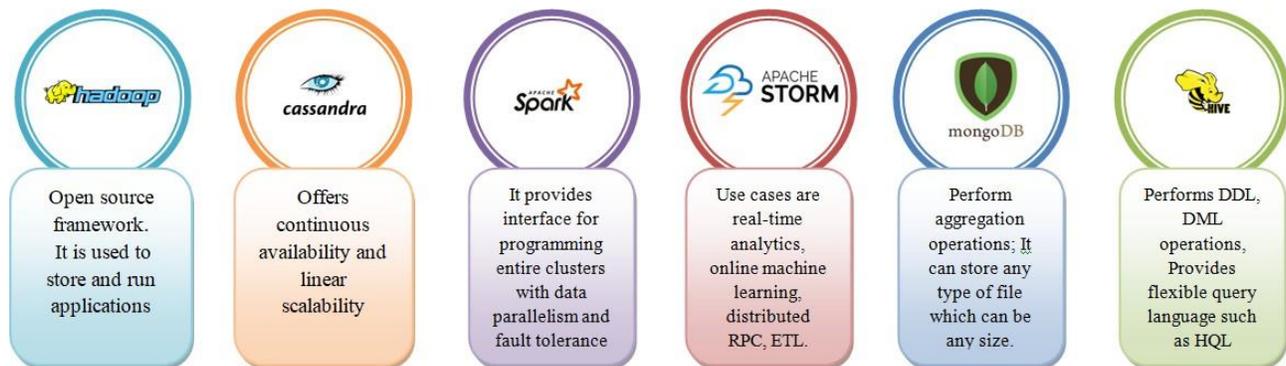

Fig 2.1: Big data tools

1. **HPCC:** Developed by LexisNexis Risk Solution. It performs on single platform single programming language and single architecture for data processing. Offers higher redundancy, availability, performance and scalability. Optimizes code for parallel processing.

2. **Qubole:** Self organized and managed tool. It's a single platform for every use case, comprehensive security, and compliance. Provide actionable alerts, recommendations, insights, open source engines and optimized for the cloud.

3. **Statwing:** Explore any form of data in seconds. Helps in explore relationships, create charts in minutes and clean data. Helps in creating histograms, heat maps, bar chart in excel or PowerPoint, scatter plots and heat maps.

4. **Couch DB:** Stores JSON document can be accessed using web or query in JavaScript. It provide fault tolerant storage and distributed scaling. Accessing of data is defined by Couch replication protocol. Couch DB is a

International Research Journal of Engineering and Technology (IRJET)   e-ISSN: 2395-0056

Volume: 07, Special Issue | June 2020   **www.irjet.net**   **p-ISSN: 2395-0072**

International Conference on Recent Trends in Science & Technology-2020 (ICRTST - 2020)

Organised by: ATME College of Engineering, Mysuru, INDIA

single-node database.

5. **Flink:** It's a open source stream processing tool in big data. It is accurate streaming application and provides High performance, availability, distributed, fault tolerant, good throughput, latency characteristics, supports third party system for data source and sinks.

6. **Cloudera:** Provide provision for multi cloud, Develop and train data models, deliver real time insights for monitoring and detection. Deploy and manage Microsoft azure, google cloud platform, cloudera enterprises across AWS.

7. **Kaggle:** It is the largest data community. Helps researchers and organizations to post their data and statics. Contribute to the open data movement and best platform to analyze open data seamlessly.

3. BIG DATA IN REAL TIME

Big data lies how organization is using collected data not on how much data has been getting collected. There is lot of big data solutions that makes big data analysis efficient and easy. These solutions are used to obtain benefits from increasing level of data in all industry verticals. Here some of the applications of big data are in discussed [7].
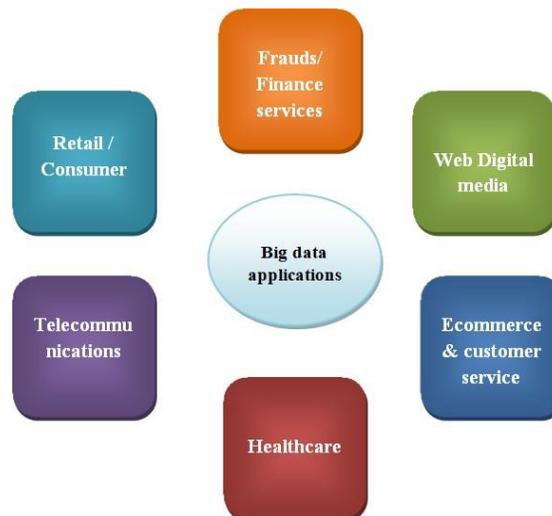

Fig 3.1: Applications of big data

1. **Healthcare**: No need to carry out unnecessary diagnosis and treatment cost is very less. It helps to avoid preventable diseases detecting in beginning stage. Eg Sensors and wearable devices have been introduced.

2. **Government Sector**: Governments are of any country comes with huge amount of data in every day basis. Big data are used in Cyber security: To catch hold tax evaders, Welfare schemes: to overcome national challenges such as terrorism, energy resource exploration, unemployment. Eg Food Drug Administration (FDA) works on

the analysis of big data.

3. **Entertainment Industry and Media:** People with various accesses to digital gadgets usage made data generation more. Entertainment and industry is the main cause for huge data. Big data predicts interest of the users. Get feedback from the customers and targeting advertisements effectively. Eg Amazon Prime.

4. **Weather Patters:** Big data using satellite and sensors used to study global warming, weather forecasting and make required preparation to avoid natural disaster it also predict amount of usable water available around the world. Eg IBM Deep thunder: research project by IBM helps in weather forecasting through high performance computing of big data. Tokyo from IBM predicting the probability of damaged power lines.

5. **Transportation Industry:** Big data contributes in congestion management and traffic control, safety level of traffic, route planning. Eg Uber gives information about vehicle, driver, fare charge for trip etc it also predicts best route to save time and fuel. Google maps to locate least traffic routes.

6. **Big Data in banking sector:** Big data helps to track illegal activities such as: Misuse of debit/ credit cards also works on business clarity, risk mitigation, money laundering, and customer statistics alteration. Eg Anti money laundering software SAS AML to detect suspicious transactions and analyzing customer data.

7. **Education Industry:** Education sector use big data in Customized and dynamic learning programs, grading systems, career prediction, reframing course material.

4. CHALLENGES OF BIG DATA

1. **Dealing with data growth:** Storing and analyzing all data is difficult task. Most of the data are unstructured data such data and photos, text documents, audio, videos are difficult to search and analyze.[8][9][10]
2. **Generating insights in timely manner:** Creating new innovations, launching new product and services, decreasing expenses through operational cost effectiveness.
3. **Validating data:** It is important factor to identify and improve big data quality. Big data quality is designed to study original big data, data quality its dimension, data validation and its tools.
4. **Securing big data:** It is one of the big concerns in all the organization. Data security methods should be improved and should also use additional features such as data segregation (42%), data encryption (52%) and access control (59%).[11][12][13]
5. **Integration of big data:** Data comes from different places such as media, email systems, enterprise applications and many other sources aggregating data and reconciling data is difficult task. [14]

5. FUTURE SCOPE

The amount of data generation will be more in future so experts are needed to handle such complex data so more employment opportunities will be open. New policies for privacy will be made; it paves the way for safer ecosystem for consumers to generate data. Redundancy will be avoided by big data analytics eg HR analytics upskilling data analytics in recruitment, budget, retrenchment, appraisals, and wage differences. Dark data will make the way for future big data.

6. CONCLUSION

Future of big data its analytics and maintenance is intensive. Data is created with fast velocity and with large variety and quality of this data needs to be processed. Some data are processed with batch processing and some with real time analytics. Researchers can solve problems of big data efficiently by using different techniques such as data mining, machine learning, cloud computing, data stream processing and quantum computing. This paper discuss on Big data applications, tools, challenges and its future scope.

References

[1]   https://www.softwaretestinghelp.com/big-data-tools [Accessed on 17 th January 2020]

[2]   Dr. Venkatesh Naganathan, Comparative Analysis of Big Data, Big Data Analytics: Challenges and Trends, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 05, May-2018

[3]   https://towardsdatascience.com/8-open-source-big-data-tools-to-use-in-2018 [Accessed on 17 th Jan 2020].

[4]   Kavyashree E.D1, Gagana M.S2, Vartika Sharma3, SURVEY ON SECURITY ISSUES AND CHALLENGES ON MAPREDUCE IN BIG DATA ANALYTICS, 3 rd National Conference on Emerging Trends in Electronics and Communication (NCETEC-16) 2016.

[5]   Vidyashree H D, Kavyashree E D, Sowmya Shree P, A Review on Big Data- Storage Techniques and Its Challenges, 2018 IJSRST, Volume 4, Issue 5, April 2018.

[6]   https://big-data-tools.com [Accessed on 21 January 2020].

[7]   https://www.greycampus.com/opencampus/big-data-developer/applications-of-big-data[Accessed on 20 Jan 2020]

[8]   Samiddha Mukherjee1, Ravi Shaw2, Big Data – Concepts, Applications, Challenges and Future Scope, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016.

[9]   D. P. Acharjya, Kauser Ahmed P, A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 7, No. 2, 2016.

[10] http://mivory.dbsdataprojects.com [Accessed on 20th January 2020]

[11] https://intellipaat.com[Accessed on 21 January 2020]

[12] Jens Baum 1, Christoph Laroque 1,* , Benjamin Oeser 1, Anders Skoogh 2 andMukund Subramaniyan, Applications of Big Data analytics and Related Technologies in Maintenance—Literature-Based Research,mdi.com[Accessed on 21 January 2020]

[13] https://www.datamation.com[Accessed on 21 January 2020]

[14] Althaf Rahaman.Sk1, Sai Rajesh.K2, .Girija Rani K3, Challenging tools on Research Issues in Big Data Analytics, IJEDR, Volume 6, Issue 1, 2018.