ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022

BIG DATA-A LITERATURE REVIEW

Kavita Divekar[1], Priya Mathurkar[2], Dr.Mahima Singh[3]
Pratibha Institute of Business Management, Chinchwad, Pune, India

ABSTRACT:
In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. It commences with the concept of the subject in itself along with its characteristics and the two general approaches of dealing with it. The comprehensive study further goes on to elucidate the applications of Big Data in all diverse aspects of economy and being. Beside this, the incorporation of Big Data in order to improve population health, for the betterment of finance, telecom industry, food industry and for fraud detection and sentiment analysis have been delineated and different trends in big data is also discussed.
Keywords: Big Data, 4V's, Sentiment Analysis, Trends

INTRODUCTION

Online Shopping Behavior

Every day, we create 2.5 quintillion bytes of data so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals, such colossal amount of data that is being produced continuously is what can be coined as Big Data. Big Data decodes previously untouched data to derive new insight that gets integrated into business operations. However, as the amounts of data increases exponential, the current techniques are becoming obsolete. Dealing with Big Data requires comprehensive coding skills, domain knowledge and statistics. Despite being Herculean in nature, Big Data applications are almost ubiquitous- from marketing to scientific research to customer interests and so on. We can witness Big Data in action almost everywhere today. From Facebook which handles over 40 billion photos from its user base to CERN's Large Hydron Collider (LHC) which generates 15PB a year to Walmart which handles more than 1 billion customer transactions in an hour.

CHARACTERISTICS OF BIG DATA

Big data is data whose scale, distribution, diversity, and/or timeliness require the use of new technical architectures, analytics, and tools in order to enable insights that unlock new sources of business value. Four main features characterize big data:

Volume:

Volume is the primary attribute of big data. Big data can be quantified by size in TBs or PBs, as well as even the number of records, transactions, tables, or files of the data is its size, and how enormous it is.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022**

Velocity:

Velocity refers to the rate with which data is changing, or how often it is created. This is basically the frequency of data generation or the frequency of data delivery.

Variety:

Documents to databases to excel tables to pictures and videos and audios in hundreds of formats, data is now losing structure. Structure can no longer be imposed like before for the analysis of data. Data generated can be o any type- structures, semi-structured or unstructured.

Veracity:

Veracity focuses on the quality of the data. This characterizes big data quality as good, bad, or undefined due to data inconsistency, incompleteness, ambiguity, latency, deception, and approximations.

There are at present two general approaches to big data:

a. **Divide and Conquer using Hadoop:** The huge data set is broken into smaller parts and processed in a parallel fashion using many servers.

b. **Brute Force using technology on the likes of SAP HANA:** One very powerful server with massive storage is used to compress the data set into a single unit.

APPLICATIONS

Big Data is a field which can be used in any zone whatsoever given that this large quantity of data can be harnessed to one‘s advantage. The major applications of Big Data have been listed below.

I. The Third Eye- Data Visualization

Organizations worldwide are slowly and perpetually recognizing the importance of big data analytics. From predicting customer purchasing behavior patterns to influencing them to make purchases to detecting fraud and misuse which until very recently used to be an incomprehensible task for most companies big data analytics is a one-stop solution. Business experts should have the opportunity to question and interpret data according to their business requirements irrespective of the complexity and volume of the data. In order to achieve this requirement, data scientists need to efficiently visualize and present this data in a comprehensible manner. Giants like Google, Facebook, Twitter, EBay, Wal-Mart etc., adopted data visualization to ease complexity of handling data. Data visualization has shown immense positive outcomes in such business organizations. Implementing data analytics and data visualization, enterprises can finally begin to tap into the immense potential that Big data possesses and ensure greater return on investments and business stability.

II. Integration- An exigency of the 21st century

Integrating digital capabilities in decision-making of an organization is transforming enterprises. By transforming the processes, such companies are developing agility, flexibility and precision that enables new growth. Gartner described the confluence of mobile devices, social networks, cloud services and big data analytics as the as nexus of forces. Using social and mobile technologies to alter the way people connect and interact with the organizations

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022**

and incorporating big data analytics in this process is proving to be a boon for organizations implementing it. Using this concept, enterprises are finding ways to leverage the data better either to increase revenues or to cut costs even if most of it is still focused on customer-centric outcomes. Such customer-centric objectives may still be the primary concern of most companies, a gradual shift to integrating big data technologies into the background operations and internal processes.

III. Big Data in Healthcare:

Healthcare is one of those arenas in which Big Data ought to have the maximum social impact. Right from the diagnosis of potential health hazards in an individual to complex medical research, big data is present in all aspects of it . Devices such as the Fitbit , Jawbone and the Samsung Gear Fit allow the user to track and upload data. Soon enough such data will be compiled and made available to doctors, which will aid them in the diagnosis. Several partnerships like the Pittsburgh Health Data Alliance have been established. The Pittsburgh Health Data Alliance is a collaboration of the Carnegie Mellon University, University of Pittsburgh and the UPMC. The health care field generates an enormous amount of data every day. There is a need, and opportunity, to mine this data and provide it to the medical researchers and practitioners who can put it to work in real life, to benefit real people. The solutions we develop will be focused on preventing the onset of disease, improving diagnosis and enhancing quality of care. Further, there is the potential to lower health care costs, one of the greatest challenges facing our nation. Big Data solutions can help the industry acquire organize & analyze this data to optimize resource allocation, plug inefficiencies, reduce cost of treatment, improve access to healthcare & advance medicinal research.

IV. Big Data and the World of Finance:

Big Data can be a very useful tool in analyzing the incredibly complex stock market moves and aid in making global financial decisions. For example, intelligent and extensive analysis of the big data available on Google Trends can aid in forecasting the stock market. Though this is not a fool-proof method, it definitely is an advancement in the field. A research study by the Warwick Business School drew on records from Google, Wikipedia and Amazon Mechanical Trunk in the time period of 2004-2012 and analyzed the link between Internet searches on politics or business and stock market moves.

V. Big Data in Fraud Detection:

Forensic Data Analytics or FDA has been an intriguingarea of interest in the past decade. However, very few companies are actually using FDA to mine big data. The reasons for this unfortunate situation vary from the deficit of expertise and awareness, developing the right tools to mine big data to lack of appropriate technology and inability to handle such humungous quantities of data. Ernst & Young undertook the Global forensic data analytics survey in 2014 and found that, ―Our survey finds that 42% of companies with revenues between US$100 million to US$1 billion are reviewing less than 10,000 records. And 71% companies with more than US$1 billion in sales report examining just one million records or fewer. Companies know there are high risk numbers in book entries, such as round thousands

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022**

or duplicates, but they're only just starting to analyze descriptions for those book entries. Looking at both the numbers and words can mean the difference between uncovering fraud, and falling victim to it.The combination of appropriate data and big data analytics can help combat fraudulent activities. Though several companies are mining big data for this purpose there are still limitations in their approach. They are either keeping the data siloed, limiting the analysis to be performed or only taking into consideration the structured data thus only giving a subset of information. A more holistic approach to the implementation of big data analytics is required. Companies such as Pactera is developing solutions which will process massive amounts of structured and unstructured data and develop varied models and algorithms to find patterns of fraud and anomalies and predict customer behavior.

VI. Big Data and the Food Industry:

The impact of Big Data on the food industry is increasing exponentially. Be it for tracking the quality of products or presenting recommendations to the customer or developing marketing strategies for better customer experience, the presence of Big Data analytics on the food industry is slowly becoming ubiquitous. IBM collaborated with The Cheesecake Factory to analyze structured data like restaurant's location and unstructured data such as flavours to increase customer satisfaction. In a news article, it stated, ―N2N has teamed up with IBM to provide The Cheesecake Factory with a technology that can communicate critical supply chain data instantly, so thousands of food items won't need to be recalled and tested. Nardone said they have initiated a conversation with the Centers for Disease Control and Prevention, as it may be easier to track the culprit if a food-related scandal occurs. Similarly, apps such as the Food Genius applies big data to predict specific recommendations to the customers. The company accumulates menu-level data parsed with ingredients, preparation methods, spices etc. and then analyzes them with individual customer preferences to determine trends and aid food giants make marketing strategies. Companies such as Starbucks, Dominos and Subway take advantage of big data analytics to track individual customer preferences and present customers with personalized offers so as to increase customer base and improve customer satisfaction

VII. Big Data for the Telecom Industry:

In order to improve customer service and satisfaction, concepts of Big Data and Machine Learning are being progressively implemented. Call detail records, web and customer service logs, emails to social media as well as geospatial and weather data are the few examples of data being accessible to telecom operators. Handling such massive amounts of data can be a daunting task. Developing deep insights with the aid of Machine Language running on Apache Hadoop can help operators to economically take advantage of the ever-increasing datasets so as to enhance their quality of service and customer experience as well as to increase the customer base with ad targeting and promotions and reduce the operational costs. The benefits of using such technologies are immense. Predictive maintenance ensures that operational disruptions are predicted, prevented and recovered. Real-time processed data can be used to dynamically allocate the bandwidth to reduce congestion

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022

TOP 10 BIG DATA TRENDS FOR 2017

I. The Proliferation of Big Data

Proliferation of big data has made it crucial to analyze data quickly to gain valuable insight. Organizations must turn the terabytes of big data that is not being used, classified as dark data, into useable data.Big data has not yet yielded the substantial results that organizations require to develop new insights for new, innovative offerings to derive a competitive advantage

II. The Use of Big Data to Improve CX

Using big data to improve CX by moving from legacy to vendor systems, during M&A, and with core system upgrades. Analyzing data with self-service flexibility to quickly harness insights about leading trends, along with competitive insight into new customer acquisition growth opportunities. Using big data to better understand customers in order to improve top line revenue through cross-sell/upsell or remove risk of lost revenue by reducing churn.

III. Wider Adoption of Hadoop

More and more organizations will be adopting Hadoop and other big data stores, in turn, vendors will rapidly introduce new, innovative Hadoop solutions. With Hadoop in place, organizations will be able to crunch large amounts of data using advanced analytics to find nuggets of valuable information for making profitable decisions.

IV. Hello to Predictive Analytics

Precisely predict future behaviors and events to improve profitability.Make a leap in improving fraud detection rapidly to minimize revenue risk exposure and improve operational excellence.

V. More Focus on Cloud-Based Data Analytics

Moving data analytics to the cloud accelerates adoption of the latest capabilities to turn data into action.Cut costs in ongoing maintenance and operations by moving data analytics to the cloud.

VI. The Move toward Informatics and the Ability to Identify the Value of Data

Use informatics to help integrate the collection, analysis and visualization of complex data to derive revenue and efficiency value from that data.

VII. Achieving Maximum Business Intelligence with Data Virtualization

Data virtualization unlocks what is hidden within large data sets. Graphic data virtualization allows organizations to retrieve and manipulate data on the fly regardless of how the data is formatted or where it is located.

VIII. Convergence of IoT, the Cloud, Big Data, and Cybersecurity

The convergence of data management technologies such as data quality, data preparation, data analytics, data integration and more. As we continue to become more reliant on smart devices, inter-connectivity and machine learning will become even more important to protect these assets from cyber security threats.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022**

IX. Improving Digital Channel Optimization and the Omnichannel Experience

Delivering the balance of traditional channels with digital channels to connect with the customer in their preferred channel. Continuously looking for innovative ways to enhance CX across channels to achieve a competitive advantage.

X. Self-Service Data Preparation and Analytics to Improve Efficiency

Self-service data preparation tools boost time to value enabling organizations to prepare data regardless of the type of data, whether structured, semi-structured or unstructured. Decreased reliance on development teams to massage the data by introducing more self-service capabilities to give power to the user and, in turn, improve operational efficiency.

CASE STUDY

McCain Foods Limited is an international leader in the frozen food industry and the world's largest manufacturer of frozen potato specialities, employing approximately 18,000 people and operating 50 production facilties on six continents. McCain Foods is a global food giant, with great pizzas, vegetables, appetizers, and desserts. A privately-owned company based in Canada, McCain generates annual sales in excess of CDN $6 billion.

Solution In McCain Foods

McCain sat down with CIO Roman Coba to talk about integrating data and how the integration has changed their business from the factory to the boardroom. McCain has taken more than 22.000 reports and 3,000 personal reporting systems and put the data in one place. For McCain it truly was a game changer.Innovation at McCain is all about developing new processes and new products. Teradata plays in both spaces. In retail, Teradata provides insights into what consumers want, how, where and why they spend their money. Those insights allow McCain to develop the best new product in manufacturing. Teradata provides insights in their manufacturing processes and operations, allowing McCain to optimize runtimes, equipment, and supply chain.

Conclusion

With new insight from the data, McCain transformed from a business culture used to treat data 'one way' into a culture that is asking for more data in order to innovate and create more change. Executives now look at their dashboards everyday and answer their own questions within minutes…they are able to drill down to get details and answer to solve the business problems in a day rather than weeks.

FUTURE SCOPE AND DEVELOPMENT

Today, Big Data is influencing IT industry like few technologies have done before. The massive data generated from sensor-enabled machines, mobile devices, cloud computing, social media, satellites help different organizations improve their decision making and take their business to another level. "Big data absolutely has the potential to change the way governments, organizations, and academic institutions conduct business and make discoveries, and its likely to change how everyone lives their day-to-day lives," - Susan Hauser, corporate vice president of Microsoft. Data is the biggest thing to hit the industry since PC was invented by Steve Jobs. As mentioned earlier in this paper, every day data is

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022**

generated in such a rapid manner that, traditional database and other data storing system will gradually give up in storing, retrieving, and finding relationships among data. Big data technologies have addressed the problems related to this new big data revolution through the use of commodity hardware and distribution. Companies like Google, Yahoo!, General Electric, Cornerstone, Microsoft, Kaggle, Facebook, Amazon that are investing a lot in Big Data research and projects. IDC estimated the value of Big Data market to be ―about $ 6.8 billion in 2012 growing almost 40 percent every year to $17 billion by 2015. By 2017, Wikibon's Jeff Kelly predicts the Big Data market will top $50 billion. Demand is so hot for solutions that all companies are exploring big data strategies. The problem is that the companies lack internal expertise and best practices.. the side effect is that there is a services and consulting boom in big data. It's a perfect storm of product and services‖ says Wikibon's Jeff Kelly. Recently it was announced that, Indian Prime Minister's office is using Big Data analytics to understand Indian citizen's sentiments and ideas through crowd sourcing platform www.mygov.in and social media to get a picture of common people's thought and opinion on government actions. Google is launching the Google Cloud Platform, which provides developers to develop a range of products from simple websites to complex applications. It enables users to launch virtual machines, store huge amount of data online, and plenty of other things . Basically, it will be an one stop platform for cloud based applications, online gaming, mobile applications, etc. All these required huge amount of data processing where Big Data plays an immense role in data processing. The predictions from the IDC Future Scope for Big Data and Analytics are:

1. Visual data discovery tools will be growing 2.5 times faster than rest of the Business Intelligence (BI) market. By 2018, investing in this enabler of end-user self-service will become a requirement for all enterprises.
2. Over the next five years spending on cloud-based Big Data and analytics (BDA) solutions will grow three times faster than spending for on-premise solutions. Hybrid on/off premise deployments will become a requirement.
3. Shortage of skilled staff will persist. In the U.S. alone there will be 181,000 deep analytics roles in 2018 and five times that many positions requiring related skills in data management and interpretation.
4. By 2017 unified data platform architecture will become the foundation of BDA strategy. The unification will occur across information management, analysis, and search technology.
5. Growth in applications incorporating advanced and predictive analytics, including machine learning, will accelerate in 2015. These apps will grow 65% faster than apps without predictive functionality.
6. 70% of large organizations already purchase external data and 100% will do so by 2019. In parallel more organizations will begin to monetize their data by selling them or providing value-added content.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022**

7. Adoption of technology to continuously analyze streams of events will accelerate in 2015 as it is applied to Internet of Things (IoT) analytics, which is expected to grow at a five-year compound annual growth rate (CAGR) of 30%.
8. Decision management platforms will expand at a CAGR of 60% through 2019 in response to the need for greater consistency in decision making and decision making process knowledge retention.
9. Rich media (video, audio, image) analytics will at least triple in 2015 and emerge as the key driver for BDA technology investment.
10. By 2018 half of all consumers will interact with services based on cognitive computing on a regular basis.

Big data isn't new, but now has reached critical mass as people digitize their lives. "People are walking sensors," said Nicholas Skytland, project manager at NASA within the Human Adaptation and Countermeasures Division of the Space Life Sciences Directorate . Taking an average of all the figures suggested by leading big data market analyst and research firms, it can be concluded that approximately 15 percent of all IT organizations will move to cloud-based service platforms, and between 2015 and 2021, this service market is expected to grow about 35 percent.

CONCLUSION

This literature survey discusses Big Data from its infancy until its current state. It elaborates on the concepts of big data followed by the characteristics, applications, different trends in big data and finally I have discussed a case study and opportunities that could be harnessed in this field. Big Data is an evolving field, where much of the research is yet to be done. Big data at present, is handled by the software named Hadoop. However, the proliferating amounts of data is making Hadoop insufficient. To harness the potential of Big Data completely in the future, extensive research needs to be carried out and revolutionary technologies need to be developed. Summarising, Peter Sondergaard, Senior Vice President of Gartner Research famously stated, “Information is the oil of the 21st century and analytics is the combustion engine.

REFERENCES

1. https://dzone.com/articles/10-big-data-trends-for-2017
2. MarcinJedyk, MAKING BIG DATA, SMALL, Using distributed systems for processing, analysing and managing large huge data sets, Software Professional‟s Network, Cheshire Data systems Ltd.
3. S. Ghemawat, H. Gobioff, and S. Leung, ―The Google File System.‖ in ACM Symposium on Operating Systems Principles, Lake George, NY, Oct 2003, pp. 29 – 43.
4. Jefry Dean and Sanjay Ghemwat, MapReduce:A Flexible Data Processing Tool, Communications of the ACM, Volume 53, Issuse.1,January 2010, pp 72-77.