

Applications, Challenges and Future Scope - Big Data

[¹]D. Rajkumar, [²] Poornachander. V, [³] Marapaka Nagaraju

[¹] Department of Computer Science, Kakatiya Government College, Hanamkonda, Warangal (Urban), Telangana State

[²] Research Scholar, Department of Computer Science, Osmania University, Hyderabad, Telangana State.

[³] Research Scholar, Department of Computer Science, Osmania University, Hyderabad, Telangana State

Abstract- The phrase Big Data has been coined to submit to the gargantuan largeness of data that cannot be dealt with by traditional data-handling techniques. Big Data is still a novel concept and in the following literature we determined to elaborate it in a conspicuous fashion. It commences with the concept of the subject itself along with its properties 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. The exploitation of Big Data Analytics after integrating it with digital capabilities to secure business growth and its visualization to make it understandable to the technically apprenticed business analyzers has been discussed in depth. Aside 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. The challenges that are encumbering the growth of Big Data Analytics are accounted for in depth in the paper. This topic has been isolated into two arenas- one being the practical challenges faces whilst and other being the theoretical challenges. The hurdle of securing the data and democratizing it have been elaborated amongst several others such as inability in finding sound data professionals in required amounts and software that possess ability to process data at a high rate. Through the article, the authors intend to decode the notions in an intelligible manner embodying in text several use-cases and illustrations.

Keywords: Big Data, Data Visualization, Integration, Encryption, Data Democratization.

I. INTRODUCTION

We generate 2.5 quintillion bytes of data in every day, 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 are used to gather posts to social media sites, digital pictures and videos, climate information, purchase transaction records, and cell phone GPS signals to name a few. Such vast 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 approaching that gets integrated into business operations. However, 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 universal from marketing to scientific research to customer interests and so on. We can witness Big Data in action almost everywhere today. Big Data can be simply defined by explaining the 3V's volume, velocity and variety which are the driving dimensions of Big Data quantification. Gartner analyst, Doug Laney introduced the famous 3 V's concept in his 2001 Metagroup publication. 3D data management: It controls Data Volume, Variety and Velocity.



Image-1: schematic representation of the 3V's of Big Data

a. Volume: Essentially it concerns the large quantities of data that is generated continuously. Initially storing such data was problematic because of high storage costs. However this is only a better technology and temporary solution needs to be developed. E-Commerce, Smartphone and social networking websites are examples where enormous amounts of data are being generated. This data can be easily distinguished between structured data, unstructured data and semi-structured data.

b. Velocity: This technique is only realistic when the incoming data rate is slower than the batch processing rate and the delay is much of a hindrance. At present times, the speed at which such huge amounts of data are being generated is unbelievably high.

c. Variety: Data generated can be any type i.e., structures, semi-structured or unstructured. The conventional form of data is structured data. For example unstructured data can be generated from social networking sites, sensors and satellites. Implementing Big Data is a huge task given the large volume, velocity and variety.

Big Data is a term encircling the use of techniques to visualize, process, capture, and analyze potentially large datasets in a reasonable timeframe not accessible to standard IT technologies. By expansion, the platform, tools and software are used for this purpose are together called Big Data technologies.

Currently, the most commonly implemented technology is Hadoop. Hadoop is the conclusion of several other technologies like Hadoop Distribution File Systems, Pig, Hive and HBase. Etc. However, even Hadoop or other existing techniques will be highly incapable of dealing with the complexities of Big Data in the near future.

The following are few cases where standard processing approaches to problems will fail due to Big Data

- Large Synoptic Survey Telescope (LSST): Every night during the decade –long LSST survey sky over 30 thousands gigabytes (30TB) of images will be generated There is a corollary to Parkinson’s Law that states.
- Data expands to fill the space available for objectives storage. This is no longer true since the data being generated technologies into the background operations and internal will soon exceed all available storage space.
- 72 hours of video are uploaded to You Tube every minute.

At present there are two general approaches in big data

a. Divide and Conquer using Hadoop: The huge data set is broken into smaller parts and process 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.

II. APPLICATIONS

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

The Third Eye- Data Visualization

In order to get this constraint, data scientists need to efficiently visualize and present this data in comprehensible manner. Giants like Google, Face book, 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 finally begin to tap into the immense potential that Big data possesses and ensure greater return on investments and business stability.

Integration- Integrating digital capabilities in decision-making of an organization is transforming enterprises is an exigency of the 21st century. By transforming the processes, such companies are developing agility, flexibility and precision that enables new growth.

Big Data in Healthcare:

Every day health care field generates an enormous amount of data. There is a need and chance 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. And the Alliance will also drive economic growth in Pittsburgh, attracting hundreds of companies and entrepreneurs, and generating thousands of jobs, from around the world.

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 Big Data is also being implemented in a field called Quantitative Investing where data scientists with negligible financial training are trying to incorporate computing power into predicting securities prices by drawing ideas from sources like newswires, earning reports, weather bulletins, Facebook and Twitter.

In general, big data is set to develop the landscape of Finance and Economy. Several financial institutions are adopting big data policies in order to gain a spirited edge. Complex algorithms are being developed to execute trades through all the structured and unstructured data gained from the sources. The methods adopted so far has not been completely adept, however, extensive research ensures growing dependence of the stock markets, financial organizations and economies on big data analytics.

Big Data and Sentiment Analysis:

Sentiment Analysis is by far the most comprehensively used application of big data. Presently, zillions of conversations are occurring on the social media, which when harness to one’s advantage can aid any company in determining new

patterns, protecting their brand image and segmenting consumer base to improve product marketing and the overall customer experience. Several giants are presently developing tools for efficient sentiment analysis.

Big Data for the Telecom Industry

Using concepts of Big Data and Machine learning are being progressively implemented to improve the customer service and satisfaction. 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.

Challenges in Big Data:

- One of the key set of challenges faced in today's tight market is need to find and analyze the required data at the least speed possible. However with exponentially growing amount of data, speed becomes a major issue as analyzing such absolute volumes of data in detail to find out required output becomes more and more tedious. It is not only the quantity of data but also discovering the data according to the appropriateness of the project which is a Herculean task.
- Elimination of out-of-context data is an essential objective. Even if in-context data retrieved at a high speed is achieved, the quality of data may be compromised if it is not accurate or timely.
- It involves those relating to the vulnerability and security of Big Data. Breaches of privacy, especially with data relating to individuals and organizations have been a topic of serious concern.
- Organizations dealing with big data need to take this issue in their stride and make sure that the data storage and location be made heavily protected so that it is not misused. They could do so by using unique database tables, having dedicated database servers, encrypting the data, having multiple security levels, having separate authentication and authorization modules and ensuring secure system operations, data transmission and data flow control.

Three key areas of security threats have been identified in the implementation of Big Data using software such as Hadoop.

- Breach of privacy by unauthorized release of data,
- Manipulation of data in the database and
- Denial of information.

In particular, the following areas of threat have been recognized in Hadoop.

- Manipulation of data in a file at a DataNode through pipeline-streaming data-transfer protocol.
- Adding/deleting/changing priority of a job in a queue.

- Unauthorized access of intermediate data of Map job via its task trackers HTTP shuffle protocol.
- An executing task may use the host operating system interfaces to access other tasks, access local data which include intermediate Map output or the local storage of the DataNode that runs on the same physical node.
- Submitting a workflow to Oozie as another user.
 - Big Data analysts is faced the Real time security or compliance monitoring is also a challenge. Due to the copious amounts of data involved, the number of alarms triggered by the security devices is so large that several of these alarms tend to be overlooked as humans cannot cope with the shear amount.

The above challenges that are faced by Big Data needs to be addressed and solutions of these problems need to be determined so that industries can start implementing big data analytics in their business strategies

FUTURE SCOPE AND DEVELOPMENT

Future Scope for Big Data and Analytics are:

1. Visual data discovery tools will be growing 2.5 times quicker than rest of the Business Intelligence (BI) market. By 2018, invest in this enabler of end user self-service will become a requirement for all enterprises.
2. In excess of next five years spending on cloud-based Big Data and analytics (BDA) solutions will grow three times quicker than spending for on-premise solutions.
3. Hybrid on/off premise deployments will become arequirement.
4. Five times that many positions requiring related Skills there will be 181,000 deep analytics roles in 2018 and Shortage of skilled staff will persist.
5. Growth in applications incorporating advanced and analytics include machine learning will speed up in 2015. These apps will grow 65% faster than apps without predictive functionality.
6. 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 makes. process knowledge retention.

V. CONCLUSION

A survey of this literature discusses Big Data from its infancy until its current state. It elaborates on the concepts of big data followed by the applications and the challenges faced by it. Finally we have discussed the future

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 4, April 2018

opportunities that could be connect in this field. Big Data is an evolving field, where much of the research is yet to be done. Present Big data is handled by the software named Hadoop. However, the proliferate amounts of data are making Hadoop insufficient. To attach the potential of Big Data completely in the future, extensive research needs to be carried out and revolutionary technologies need to be developed. Summarizing, 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. Ruchi Verma, Sathyan Ramakrishna Mani, —Using Analyrtics for PIG Tutorial YahooInc., <http://developer.yahoo.com/hadoop/tutorial/pigtutorial.html>
2. IBM-What.is.Jaql, [www.ibm.com /software /data /infosphere /hadoop /jaql /](http://www.ibm.com/software/data/infosphere/hadoop/jaql/)
3. Dryad - Microsoft Research, <http://research.microsoft.com/en-us/projects/dryad/>
4. SAS, The power to know, —Five Big Data Challenges And how to overcome them with visual analytics
5. [http://www.forbes.com /sites /gartnergroup /2013 /03 /27/ gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/](http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/)
6. [http://www.internetlivestats.com /google-search-statistics](http://www.internetlivestats.com/google-search-statistics)
7. [http://www.internetlivestats.com /twitter-statistics /](http://www.internetlivestats.com/twitter-statistics/)
8. [http://lsst.org /lst /google](http://lsst.org/lst/google)
9. [http://www.forbes.com /sites /bernardmarr /2015 /04 /21/how-big-data-is-changing-healthcare/](http://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/)