International Journal of Engineering, Science and Mathematics
Vol. 6 Issue 3, July 2017,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com

Applications of BIG DATA

Jatinder Kumar
Assistant Prof.
A. S.College, Khanna

ABSTRACT: The term Big Data has been coined to refer to the bulk of data that cannot be dealt with by traditional data-handling techniques. The concept of Big Data is still in infancy and is a novel concept.A small data becomes BIG Data when it is collected from the different resources like social media, business system, financial system, governance,banking,insurance,health care etc. In the earlier times, few companies were generating data, all others were consuming data but as per the new model, all of us are generating data, and all of us are consuming data. Now the challenge is to manage,store and process it and to use it for the decision making process.The aim of this paper is to explore the methodology in which BIG Data is can be used to provide meaningful information from the hidden patterns and hence to use it for making better fast and effective decisions to compete well in this competitive global environment.The paper throws light on various applications of Big Data in all diverse aspects of economy population health, for the betterment of finance, telecom industry, food industry and for fraud detection and sentiment analysis have been delineated.

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

INTRODUCTION

Every day, we create thousands of bytes of dataeveryday. And moreover, 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 to name a few. Such a huge 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 to extract hidden knowledge from it. The knowledge hidden in the huge databases can be used in the fields as diverse as pharmacology, finance, fraud detection, and intelligence analysis, better analysis and decision making can be facilitated by taking into consideration large amounts of heterogeneous data from many sources in many formats, and degrees of structure, and update rates. In order to address this problem, we require three things.

The need of the hour is to load and queryvery large data sets that exceed the reasonable processing capabilities of even high end server platforms. Second, those data sets are heterogeneous and interesting data often appears after the system has been deployed, so we must be able to dynamically align the schema for those data sets and to continuously integrate new data. Third, we require the ability to maintain data provenance and drill down into the source detail.

According to Gartner 2012, "Big Data are high-volume, high-velocity, and/or high- variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization".

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 used in BIG Data:

a. Volume: This essentially concerns the large quantities of data that is generated continuously. Initially storing such data was problematic because of high storage costs. However with decreasing storage costs, this problem has been kept somewhat at bay as of now. However this is only a temporary solution and better technology needs to be developed. Smartphones, E-Commerce and social networking websites are examples where massive amounts of data are being generated. This data can be easily distinguishes between structured data, unstructured data and semi-structured data.

b. Velocity: In what now seems b. Velocity: In what now seems like the pre-historic times, data was processed in batches. However this technique isonly feasible 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 colossal amounts of data are being generated is unbelievably high. For Example, Facebook generates 2.7 billion like actions/day and 300 million photos amongst others roughly amounting to 2.5 million pieces of content in each day while Google Now processes over 1.2 trillion searches per year worldwide.

c. 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. The conventional form of data is structured data. For example text. Unstructured data can be generated from social networking sites, sensors and satellites.

Implementing Big Data is a mammoth task given the large volume, velocity and variety. —Big Data is a term encompassing the use of techniques to capture, process, analyze and visualize potentially large datasets in a reasonable timeframe not accessible to standard IT technologies. By extension, the platform, tools and software used for this purpose are collectively called —Big Data technologies.Currently, the most commonly implemented technology is Hadoop. Hadoop is the culmination 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.

Now,the new challenges is front of new generation to manage the big data and find the hidden information from it. Some new methodology is used HADOOP. HADOOP-Hadoop was created by Doug Cutting, the creator of Apache Lucene, the widely used text search library. Hadoop has its origins in Apache Nutch, an open source web search engine, itself a part of the Lucene project. Today, Hadoop is a collection of related subprojects that fall under the umbrella of infrastructure for distributed computing. These projects are hosted by the Apache Soft ware Foundation, which provides support for a community of open source software projects. Although Hadoop is known for MapReduce and its distributed filesystem (HDFS, renamed from NDFS), the other subprojects provide complementary services, or build on the core to add higherlevel abstractions Core- A set of components and interfaces for distributed filesystems and general I/O (serialization, Java RPC, persistent data structures).

II. APPLICATIONS

Big Data is slowly becoming ubiquitous. Every arena of business, health or general living standards now can implement big data analytics. To put simply, 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'sadvantage. The major aplications of Big Data have been listed below:

a) 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 visualizeand 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 Bigdata possesses and ensure greater return on investments and business stability.

b) Integration- An exigency of the 21stcenturyintegrating digital capabilities in decision-making of anorganization 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 theorganizations 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.

c) Healthcare:Healthcare is one of those arenas in which Big Data oughtto have the maximum social impact. Rightfrom 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. 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 thepotential 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… The patients diagnosis will be analyzed and compared with the symptoms of others to discover patterns and ensure better treatment. IBM has taken initiative in a large scale to implement big data in healthcare systems be in its collaboration with healthcare giant Fletcher Allen or with the Premier healthcare alliance to change the way unstructured but useful clinical data is made available to more medical practitioners so as to improve population health. Big Data can also be used in major clinical trials like cure for various forms of cancer and developing tailor-made medicines for individual patients according to their genetic makeup.

d) **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 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.

A 10 step approach has been suggested by Infosys to implement analytics for fraud detection:

1. Perform SWOT analysis of existing fraud detecting paradigms.
2. Assign a dedicated fraud management team.
3. Developing or purchasing appropriate data analytics software.
4. Integrate siloed data and clear inefficiencies in the processes.
5. Establish rules relevant business obligations.
6. Determine thresholds for detection of error or discrepancies.
7. Implement predictive analysis to determine potential discrepancies and frauds.
8. Use Social Network Analysis or SNA to determine fraudulent activities.
9. Develop an integrated case management system.

10. Continue with extensive research to integrate existing systems of fraud detection with new set of techniques developed.

E) 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 and outages.

Challenges facing Big Data

Despite the extensive hype around Big Data in the industry today, very few companies have actually been able to implement the concept of Big Data. A survey published in 2013 by SAS in 2013 Big Data Survey Research Brief analyzed the reasons on why most industries are still delaying or refusing to pursue a big data strategy. It states —a little more than one-fifth of the respondents are still trying to learn more about big data, while others are still trying to understand the benefits of big data. Even though the industry has written countless articles, blogs and white papers about big data, there is still a significant contingent of data management professionals trying to understand the basics. The obstacles that limit the implementation of big data by any industry are aplenty. The ‗Big Data Talent Gapwhich distinctively exists even though a lot of research has gone into this field in the past decade is a massive issue.

VI. CONCLUSION

To avoid this big data problem data should be collected unstructured from the set of top boxes – multiple terabytes. The new methodology hadoop and all its components is very helpful. Analytics with HADOOP increases the satisfaction of customers. In the same way Hive – pull data in to Hive for interaction query and modeling. Analyze system in near real time. Big Data is here. Analysts and research organizations have made it clear that mining machine generated data is essential to future success. Embracing new technologies and techniques are always challenging, but as architects, you are expected to provide a fast, reliable path to business adoption. A quite message for all the people that everyone should involve to try the new tools to solve this problem or search the new methodology otherwise in future it may create a really very BIG problem. 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 tochange 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 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.

This literature survey discusses Big Data from its infancy until itscurrent state. It elaborates onthe concepts of big data followed bythe applications and the challenges faced by it. Finally we have discussed the future 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 are making Hadoop insufficient. To harness thepotential of Big Data completely in the future, extensive research needs to be carried out and revolutionary technologies need to be developed.

References

Puneet Singh Duggal, SanchitaPaul ,— Big Data Analysis: Challenges and Solutions , International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV

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.

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.

Jefry Dean and Sanjay Ghemwat, MapReduce:A Flexible Data Processing Tool, Communications of the ACM, Volume 53, Issuse.1,January 2010, pp 72-77.

PIGTutorial,YahooInc.,http://developer.yahoo.com/hadoop/tutorial/pigtutorial.html

IBM-What.is.Jaql, www.ibm.com/software/data/infosphere/hadoop/jaql/

Security Challenges: Dealing with too many issues ; International Journal of Recent Development in Engineering and Technology, Volume 3, Issue 2, August 2014.

David Loshin, —Addressing Five Emerging Challenges Of Big Data ; Progress Software, 2014

Rashmi N, Uma K M, Jayalakshmi K, Vinodkumar K P, —Big Data Security Challenges: Dealing with too many issues ; International Journal of Recent Development in Engineering and Technology, Volume 3, Issue 2, August 2014