International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072

A Scenario on Big Data

Rupali Yadavrao Charde[1], Triveni Yadaw[2], Anand kumar[3], Akash Sood[4], Sidharth Singh[5], Riya Sahu[6]

[1, 2, 3] *Master in Computer Science & Technology, Rajiv Gandhi Proudyogiki vishwavidyalaya, Bhopal, Madhya Pradesh, India*
[4] *NPD Engineer, Technocrat Connectivity System, Delhi, India*
[5] *Bachelor of Computer Science Engineering, Maharishi Markandeshwar University, Mullana, Haryana, India*
[6] *SONT, Rajiv Gandhi Proudyogiki vishwavidyalaya, Bhopal, Madhya Pradesh, India*

Abstract - *The proliferation in digitalized world (growth in computers memory, mobiles and storage capacity) are drawing huge attention towards the Big Data. The researchers from all part of globe from Information Technology wants to extract valuable information from science, policy, disease, market trends to help government and non-government organization to draft there policy for future. To bring such enhancement in modern world, we are ongoing on in a new concept of Big Data and almost all of businesses that are running today seek a new and better approach to remain competitive and profitable under new policy or adapts a new policy on basis of past changes . To do this, Big Data leads them in a path that stays ahead of the curves. In such consideration big data is one of the best one in all means and required proper art to be develop. Such inconceivable percentage of glean key information stored and created every minute by everyone, this focuses our idea on Big Data.*

Key Words: Big Data, User to User connection, Business compatibility, e-Science, Analytics

1. INTRODUCTION

A boomed in digital industrialization has increased in mobile network, cloud based infrastructure, great affordability of devices and daily evolution in technology arose incomprehensibly add large information, such potential stored information is termed as "Big Data". It can be stored in structured, unstructured and in hybrid (combination of structured and unstructured) [1]. But with this all, in addition it required organization to analyze for strategic decision. This gets into picture when industrial analyst Doug Laney in 2000 articulated to characterize the concept of big data as 3V's [2].

1.1. Volume: The amount of data collected from various resources including e-business transaction (Paypal, Payatm, Airtel Money etc), social media (Facebook, Twitter, Whatsapp), sensor (weather monitoring, space sensor) and machine to machine data (networking, IoT) by millions of user around the world. To study such massive data Hadoop provide great tool [2] [3] [4].

1.2. Velocity: The massive stored data need unprecedented speed with time constraint. In addition to device, it should be connected in parallel with smart sensor and metering device in real time process to keep the transparency of data [2] [3] [4].

1.3. Variety: Data comes in two or more formats, but majorly as structured data (numeric data in traditional databases) and unstructured data (like stock ticker data, email, financial transactions, audio, video and text documents) [2] [3] [4].

In addition the Big Data is classified into two more parts [4]:

1. A) Variability: Inconsistency of the data set at high velocity and in variety of data needed to be processed without hampering the information and manage the speed at peak load of data processing for example social media data demand increase in morning and evening [4] [5].

1. B) Complexity: The data coming from variety of sources make it difficult to link, cleanse, match and transfer. However, it is important to solve the complexity by creating arranged data (like different department books at library) to create relationship or linkage among the data [4] [5].

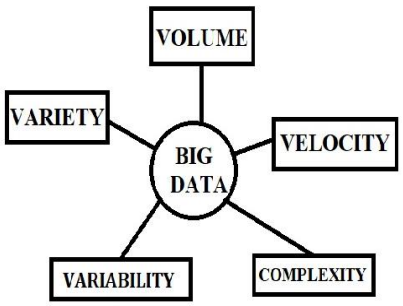

FIGURE 1: REPRESTION OF 3V'S AND OTHER KEY FEATURE OF BIG DATA

2. PRESENT BIG DATA TECHNOLOGIES

2.1 CHUKWA: Apache Chukwa is subproject of Hadoop to process large-scale data collection and analysis it. It is an open source data collecting system [6] [7]. Chukwa is used over Hadoop Distribution File System (HDFS). It's one of the advantages is replying of data to recover from errors by constructing bridge between lag handling and Map reduce [8] [9].

2.2 COLUMN-ORINTED DATABASE: A column-oriented DBMS stores data in tables form by column rather than by row. In practical it does not differ much if we store in row, in the relational DBMS world.Both columnar and row databases can use traditional database query languages like SQL to load data and perform queries. It becomes backbone of system to server, which visualizes, extract, transform and load the data [10] [11].

2.3 HADOOP: In 2006, Yahoo joined Nutch project (developed by the brain child Doug Cutting and Mike Cafarella) similar on ideas of early Google's work [12]. The Nutch has two part- the web crawler (is a Nutch) and processing and distributed computing portion become Hadoop (Nutch and Hadoop both are open source web search engine) [13] [14]. Hadoop have ability to store and process, computing power, fault tolerance, flexibility, low cost, scalability as important and significant feature [12].

2.4 HBase: Apache HBase (Apace software Foundation) provides the freedom to access the random, real-time changes (read and write) to our Big Data. Its objective is to create huge tables of billions of rows and columns to atop clusters of hardware [15]. Likewise Hadoop it is a open source, distribution, versioned, non-relational database modeled after Google [16].

2.5 HDFS: The Hadoop Distributed File System (HDFS) is designed to store very large data sets reliably, and to stream those data sets at high bandwidth to user applications. In a large cluster, thousands of servers both host directly attached storage and execute user application work. By distributing storage and computation across number of servers, the resource can grow with demand while remaining economical at every size [17].

2.6 HIVE: Apache Hive is data warehouse were all type of data being store. This is built over Apache Hadoop for providing summarization, query and analyzation. Hive stored in various database by using SQL- likewise database and the file system that integrate with Hadoop [18] [19].

2.7 SCHEME-LESS DATABASE: Scheme-less database does not required following things:

- Modeling
- Pre-setting data types in repository

But it can store data in different characteristics and can tolerate changes. It is only limited by user definition and imagination. This is also known as KV or as JSON document due to storage system [20].

2.8 STORAGE TECHNOLOGIES: At the root level its requirement is vital. For Big Data Huge storage is required to handle large data. And it must perform input/output operations per second (IOPS) necessary to deliver data to analytics tools for analyzing. In World Company like Google, Facebook, Apple, etc are the largest big data practitioners – working at hyperscale computing environment [21] [22].

3. BIG DATA PROBLEM

The rapid change in global world, changes the various scenario and prospective of our consumption and demand at various level, which bring ciaos of problem in established market. To bridge such issue a Big Data enhanced the process to enrich the system to grow in constraint environment. But still security is one of major concern, due to the fact of losing security file in such a high globalised world. Perhaps with thousand problems it helps to accumulate research data for rocketing research.

The some major key notes in Big Data Problems are described below.

3.1 BIG DATA IN COMMERCE AND BUSINESS

According to estimation, companies double in every 1.2 years across the worldwide and the volume of business data also get double [23]. It can better be understand by nearby stores like V-MART, Mega Mart, and Pantaloons. Such kind of company takes advantage of machine learning to forecast the information by exploiting the hidden data. This information also them to pricing (discount or hike), inventory and supply mechanism is balanced to have significant profit [24].

In digital era every company is working on margin scale profit which bow them to have strong forecast at stretched information of supply demand and consumer type, this is achieved only by Big Data. It also helps in transaction data which is stored in company website, but in parallel it increases the chances of fraud transaction. Cyber security is over bridged and thousands of people stuck into it [3] [25].

3.2 BIG DATA IN SOCIETY ADMINISTRATION

Country with huge population has huge responsibilities like India and China at administrative level to tackle with Big Data problems [26]. For example, the different age group people need different services as kind required good education and food, where as old one needs higher health care. Such instance data manipulate the various policy of government and iron the text to rust. All kind of information is produced by all member of the society at all level [3].

McKinesy's repot show the Big Data functionalities such as reserving information patterns and knowledge which enhances the policy regulation and sudden changes required in improving the following aspects [23] [27]:

- Productivity
- Effectiveness and
- Efficiency

3.3 BIG DATA IN SCIENTIFIC RESEARCH

Eradicate development in computer science engineering become highly data-driven. For department like astronomy, metrology, social science, bioinformatics and many more are intensely dependent on past information or data for further

research. For instance the Pluto mission carried out by NASA is collection of past navigation information which excellences the mission. Such events have data intensive work which competes with volume of information and velocity of data analyses become sophisticated [28]. Consequently, such information pushes us on large scale Big Data initiative.

The President's Council of Advisors on Science and Technology (PCAST) under Network Information Technology Research and Development (NITRD) program, we need to bridges the information between knowledge and Big Data [29]. This initiative will lays ground work as platform development, infrastructure and technique in settling complex problem in science and engineering. Finally, this will put in healthy practice and benefits the society of present and next generation [30].

This can be better understanding it by case study. For instance, a sophisticated telescope is installed at ground and space to regards with very large camera. It captures universal images in huge volume and in large quantity. The study of such data become hectic, unproductive and time consuming process [28]. To pace up such informative data, scientist and researcher utilize advance computing facilities and analysis to get clear picture of universe. From this case study it project out that Big Data is disciplines in generating data, centralized storage and analysis approaches drive the system design of Big Data [31].

4. BIG DATA TOOL TECHNIQUES

To capture the value from Big Data, development of new techniques and technology for analyzing it is required. Until know only at small scale of work is done and another required effort to achieve by boosting skills in science, mathematics, economics, statistics and other expertise's. But with multi-disciplinary gives big horizon of thinking for exploiting data intensive application. To incorporate with Big Data, we need tools (Platforms) to make sense. Today's tools concentrate on following aspects:

- Batch processing tool (Apache Hadoop) [18]
- Stream processing tool (Strom and S4) [32] [33]
- Interactive analysis (Google's Dremel and Apache Drill) [34]

4.1 BIG DATA TECHNIQUE

Big Data can be achieved by extraordinary efficient process within limited run times. Reasonably it can be applied to specific application, as used by Wal-Mart in electronic transactions using machine learning and statistical pattern by exploring behavior of large previous information [35].

The valuable data for precise decision-making is achieved by data mining, machine learning, optimization methods, statistics, social network analysis and visualization approaches [35].

4.1.1 DATA MINING

Fundamentally, data mining is extracting data and identifying patterns, applying regression and association rule learning. The principle becomes more prevalent with advent of Big Data. The present application of Data Mining is more challenging as compared to the application in traditional data mining algorithms. Taking such ciaos clustering as an example such as Fuzzy C-Mean, K-Mean and hierarchical clustering [36] [37]. This application resolves the ciaos to some extent to propose some model for enhancement of software.

These business-driven needs changed simple data salvage and statistics into more multifaceted data mining. The business problem drives test of the data that helps to build a model to describe the current information that ultimately leads to the creation of the resulting report. Below figure outlines the process.

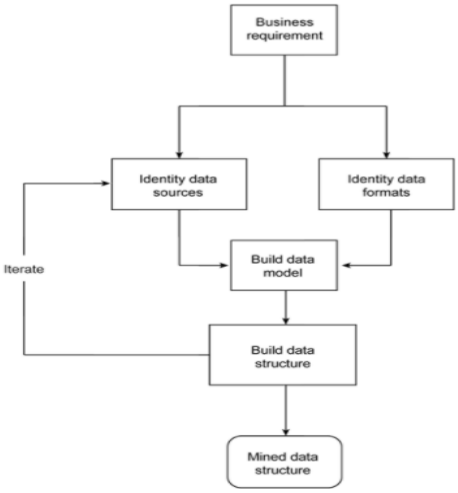

FIGURE 2: DATA MINING PROCESS [38]

4.1.2 MACHINE LEARNING

Machine learning is aimed to design to self learning capabilities that allow computers to react according to empirical data. It is subjection of artificial intelligence (AI). In concern with Big Data where user produced large data needed to be inspected and filter, an artificial intelligence (AI) is required to balance and store the information to reduce the fraud and removal of unwanted data [39].

4.1.3 OPTIMIZATION METHODS

An optimization method is applied to adequate quantitative problems, to resolve with design and economical parameter [40]. It can be applied to various field of study, such as

- Physics
- Biology
- Engineering
- Economics

However, this often creates a complexity in brain and raised the time consumption. So, real time simulation of optimization in such ciaos application is required. Data reduction and parallelization is another approach for optimization [41].

4.1.4 STATISTICS

Statistics is collection of information, organization of data and interpretation of data to study the scenario of information, which can be used to forecast the data. Statistical data are used to fetch the data by exploiting casual relationship and correlations between different objectives. Basically the data is stored in numerical form to understand the various trend and theme shown by analyzing the data.

4.1.5 SOCIAL NETWORK ANALYSIS (SNA)

It has emerged as a modern technique in modern sociology; it consists of nodes and ties. It has also gained a significant following in biology, economics, geography, history, information science, organizational studies, social psychology and development studies. SNA include social system design in behavior of human modeling [42], social networks evolution analysis [43] and graph query and mining [44].Recently, online social networks and Social media analysis have become popular.

4.1.6 VISUALIZATION APPROACHES

The basic technique used by human to sore information in his brain by visual concept, the same art is used by machine to visualize data. This data create an approach to store diagrams, images, tables and many more to understand data.

In Big Data [45], it is not easy as we do in traditional small data sets because of complexity in 3V and 4V. Visualization approach has significant impact in reducing the data size before actual rendering [46]. During data representation it's important to try and visualize the data in Big Data.

5. UNDERLYING TECHNOLOGIES AND FUTURE RESEARCHES

The ultimate aims are to promote advanced techniques and technologies developed for Big Data science is with the purpose of advancing and inventing the more sophisticated and scientific methods of managing, analyzing, visualizing, and exploiting informative knowledge from large, diverse, distributed and heterogeneous data sets. Finally benefit economic and social evolutions. A paradigm shift in scientific investigation is on the way on Big Data using granular computing, cloud computing, bio-inspired computing and quantum computing.

5.1 GRANULAR COMPUTING (GrC)

Granular computing means computing a large data in smaller size (granules) and it is a major property of GrC. The granules can be arranged in classes, cluster, groups, subsets and intervals: to build an effective computational model for complex application with gigantic (multiple granules in large size) data in information and knowledge. Hence it become very natural technique in granular computational in Big Data application. In fact GrC leads a significant transform from the current machine centric concept to human centric approach for information and knowledge. Theoretical foundation of granular involves as [36] [37] [47]

- Set theory
- Fuzzy theory
- Rough sets and
- Random sets

Su et al [48] introduced a new structure of radial basis function networks (RBFNs) that can successfully model symbolic interval-valued data. If it is done and represented in symbolic data (in Big Data) then, some algorithms and machine learning come into picture.

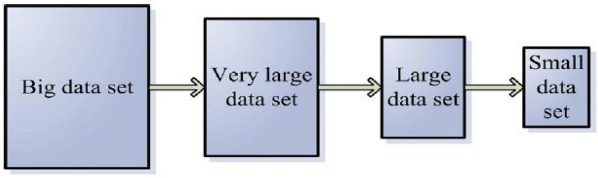

FIGURE 3: GRANULAR COMPUTING [3]

5.2 CLOUD COMPUTING

Cloud computing is the practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer. Amazon EC2 is launched in 2006 [49], since then the availability of low-cost computers, storage devices and high-capacity networks, as well as the widespread adoption of hardware virtualization, service-oriented architecture, and autonomic and utility computing has increased in cloud computing. Cloud computing is a highly feasible technology and attract a large number of researchers to develop it, and to apply on Big Data [50].

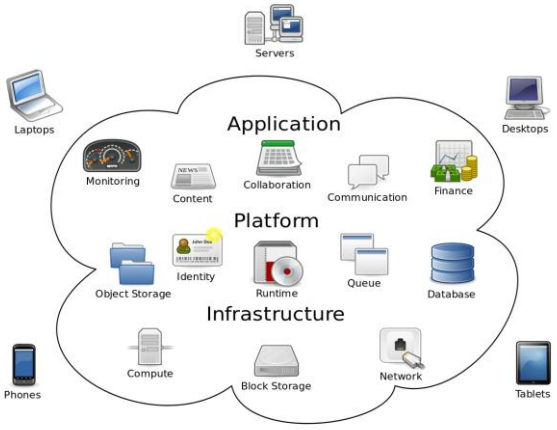

FIGURE 4: CLOUD COMPUTING [51]

5.3 BIO-INSPIRED COMPUTING

Bio-inspired computing means the biological art or way the human react to the stored data in brain (past experience- in form of Big Data) to react, perhaps in same way system react with characteristics of connectionism, social behavior and emergence[52]. Therefore, it is representing artificial intelligence based on linkage with machine learning. It relies on major subjects of biology, computer science and mathematics using it as various subsets to create Big Data.

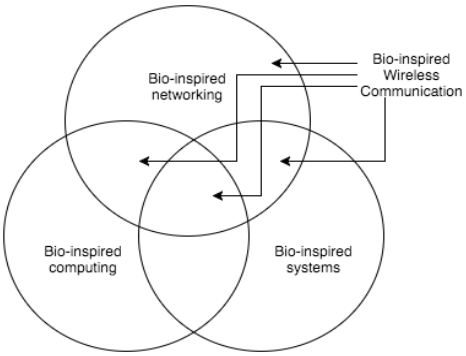

FIGURE 4: BIO-INSPIRED COMPUTING [53]

5.4 QUANTUM COMPUTING

A classical computer has a memory made up of bits, where each bit is represented by either a one or a zero. A quantum computer maintains a sequence of qubits [54].

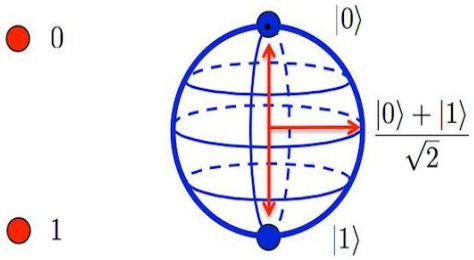

FIGURE 5: QUANTUM COMPUTING

Quantum computing is computing using quantum mechanical phenomena, such as superposition and entanglement. This also rapidly increase the computer speed by 100 times then present time computer, as we still not able to implement Big Data application on full scale due to limitation of hardware constraint which can be achieved by quantum computer.

6. APPLICATION

In today's modern world the data and information produced by machine is incomparable with regards of having best devices. But with change in technology and time the various factors limiting the application of Big Data is now not a big deals as many organization and institution come forward to enhance the market of Big Data. Some of common application at such places is as follows:-

6.1 EDUCATION

Academic institutions are now focused in finding methods for making effective learning process (by identifying scholar achievements and weakness) to help boost their skills for better future. People raised question for accountability and transparency makes it necessary to implement big data analytics in the educational institution for fair results. But not all the institution and administration are ready to accept the challenge on financial basis.

6.2 GOVERNMENT

Big Data has vast applications in the government sectors. Big data, a general term for collection the massive amount of digital data being collected from all sources, is too large, raw, or unstructured for analysis through conventional relational database techniques from all sector like

- primary sector (farming information)
- secondary sector (employed sector)and
- tertiary sector (industrial sector)

6.3 HEALTHCARE

One of the most promising areas where big data can be applied to make a change is healthcare. It has a huge potential to reduce the cost to meet the treatment (generic medicine in India is cheaper than imported medicine), predict outbreaks of epidemics, avoid preventable diseases and improve the quality of life [55]. As increased in population the health is also becoming a challenge and pushes the professional to go forward for business entrepreneurs (collecting massive amounts of data and look for best strategies to use these numbers).

6.4 INFORMATION TECHNOLOGY (IT)

The information generated by user end at various sectors such hospitals, school, government organization, social media and sharing of photo, music and video become faster and reliable with advancement in hardware in parallel with software [56]. To manage such petabyte of information large number of information technology specialist is required. To store such huge information manually at broad scale is impossible so Big Data play a vital role with IT.

6.5 INTERNATIONAL DEVELOPMENT

The increase in boundary at international level created thousand of laws (constitutional right) and policy for migration (VISA details) for exchange in alien, goods, manpower, knowledge, technology, economics and many more. To keep information of documentation and applying to large number of people it become complex and ciaos to have quick and fast service for fundamental and advance exchange. Big Data collect all kind of information and process by dividing into group and in class to make process as crystal clear and provide transparency to create any dispute at international level [57].

6.6 INTERNET OF THINGS (IoT)

Big Data and IoT works simultaneously. Extracted data gives mapping of device interconnectivity by mapping using IoT. This information generated by big data has been used by industry, government, intuition and media to forecast the audience or befitted group of people from different action. And implementation becomes easy to keep eye [58].

6.7 MANUFACTURING

Increase in customer demand and working at margin scale the forecasting of data become more important. For instance the sale of REDMI mobile on Amazon, customer gives the virtual preorder to buy [59]. Such way to predict the demand deduces the loss. This has been achieved by past cases and presents scenario Rapides the productivity according to demand.

6.8 MEDIA

Ironically lot of fraud data is generated in truth. To exploit truth lot of fraud news is broadcast on media, to filter content Big data application and security of information by cross checking by accumulating information from subsets (by data collection by mapping of information) verify the details. This all depends on sensory data collected from various source as metrological, technical, government, and scientific department regarding sudden changes at our environment to reduce loss of property and life in disaster prone area [60].

7. CONCLUSION

The integration of application and technology in digital world has been rocketed in twenty first century. The connection between user and information has changing the thoughts and understanding the world. To educate the society in right direction with security of e-information become a challenge in present society. To overcome such problems different application and software are working around as in banking, finance, e-commences and e-transaction. But still we fail to achieve to hundred percent securities. To meet such challenges the promising future is drawn from Big Data as it has capability to simulate and respond to the problem.

Big Data consist of subsets of application from storing information from granular subsets to gigantic subsets (petabyte), but limitation is the hardware and processing of the application which is resolved by using quantum technology. The scientific research has been conducted at all part of the world. Combine effort of scientist and engineer has pioneered the skills and technology to some extent to benefit some part of the society such as school, media, hospital, and government etc, perhaps lot still is needed in this direction.

Big Data has all potential to make application more comprehensive and flexible to operate as it include internet of things, cloud computing, artificial intelligence and data mining to understand for define purpose to led better future.

REFERENCES

[1] Neelam Singh, Neha Garg, Varsha Mittal "Big Data – insights, motivation and challenges"-International Journal of Scientific Research, Volume 4,Issue 12, December-2013

[2] https://www.sas.com/en_in/insights/big-data/what-is-big-data.html

[3] C.L. Philip Chen, Chun-Yang Zhang "Data-intensive application, challenges, techniques and technologies:A survey on Big Data"-ELSEVIER Information Science 275 (2014) 314-347

[4] https://www.sas.com/en_in/insights/big-data/what-is-big-data.html

[5] https://en.wikipedia.org/wiki/Big_data

[6] https://wiki.apache.org/hadoop/Chukwa

[7] http://chukwa.apache.org/

[8] https://www.thinkbiganalytics.com/leading_big_data_technologies/ingestion-and-streaming-with-storm-kafka-flume/

[9] Dr.S.Vijayrani and Ms. S. Sarmila "Research in Big Data-An Overview" – Informatics Engineering an International Journal (IEJI). Vol.4, No.3, Septemper 2016

[10] Marcus Pinnecke, David Bronske, Gabriel Compero Durand and Gunter Saake university of Magdeburg"Are Database Fit for Workload on GPU? A Storage Engine's Perspective"-International Conference On Data Engineering (ICDE) 2017.237

[11] https://en.wikipedia.org/wiki/Column-oriented_DBMS

[12] https://www.sas.com/en_in/insights/big-data/hadoop.html

[13] https://en.wikipedia.org/wiki/Apache_Hadoop

[14] http://hadoop.apache.org

[15] Nikita Bhojwani ,Asst Prof. Vatsal Shah- "A SURVEY ON HADOOP HBASE SYSTEM" International Journal of Advance Engineering and Research Development (IJAERD) Volume 3, Issue 1, January -2016, e-ISSN: 2348 - 4470, print-ISSN: 2348-6406

[16] Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach Mike Burrows, Tushar Chandra, Andrew Fikes, Robert E. Gruber-" Bigtable: A Distributed Storage System for Structured Data" Google, Inc

[17] https://www.ibm.com/analytics/hadoop/hdfs

[18] Venner, Jason (2009). Pro Hadoop. Apress. ISBN 978-1-4302-1942-2.

[19] https://en.wikipedia.org/wiki/Apache_Hive

[20] https://blog.couchbase.com/the-value-of-schema-less-databases/

[21] http://www.computerweekly.com/podcast/Big-data-storage-Defining-big-data-and-the-type-of-storage-it-needs

[22] Martin Strohbach, Jeorg Daubert, Herman Ravkin, and Mario Lischka "Big Data Storage"-New Horizons for a Data-Driven Economy, DOI 10.1007/978-3-319-21569-3_7

[23] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, Big data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, 2012

[24] http://bigdata-madesimple.com/future-of-e-commerce-5-ways-to-improve-business-with-big-data/

[25] http://dataconomy.com/2017/07/6-ways-use-big-data-ecommerce

[26] Randal E. Bryant, Data Intensive supercomputing: The Case for Disc. Technical Report CMU-CS-07-128, 2007.

[27] Roxanne Connelly, Christopher J. Playford, Vernon Gayle, Chris Dibben-" The role of administrative data in the big data revolution in social science research" Social Science Research Volume 59, September 2016, Pages 1-12

[28] Geoff Brumfiel, High-energy physics: down the petabyte highway, Nature (469) (2011) 282–283.

[29] https://insidebigdata.com/white-paper/insidebigdata-guide-to-scientific-research/

[30] https://insidebigdata.com/white-paper/insidebigdata-guide-to-scientific-research/

[31] Alessandra Aloisi, Marco Chiaberge, Gretchen Greene, Anton Koekemoer, Josh Peek, Marc Postman (team lead), Armin Rest, Daryl Swade, Jason Tumlinson, Rick White, and Brad Whitmore-" Enhancing STScI's Astronomical Data Science Capabilities over the Next Five Years"-Science Definition Team Report March 15, 2016

[32] https://gdfm.me/2013/01/02/distributed-stream-processing-showdown-s4-vs-storm

[33] Saeed Shahrivari, and Saeed Jalili –" Beyond Batch Processing: Towards Real-Time and Streaming Big Data"Computer Engineering Department, Tarbiat Modares University (TMU), Tehran, Iran

[34] https://techcrunch.com/2012/08/17/googles-real-time-big-data-tool-cloned-by-apache-drill

[35] Ms Ranju Marwaha Assistant Professor "Data Mining Techniques and Applications in Telecommunication Industry" Marwaha, International Journal of Advanced Research in Computer Science and Software Engineering 4(9), September - 2014, pp. 430-433

[36] Samarjit Das "Pattern Recognition using the Fuzzy c-means Technique" International Journal of Energy, Information and Communications Vol. 4, Issue 1, February, 2013

[37] Jin Zhou, C.L. Philip Chen, Long Chen, Hong-Xing Li, Wei Zhao, A collaborative fuzzy clustering algorithm in distributed network environments, IEEE Trans. Fuzzy Syst. PP (99) (2013) 1.

[38] https://www.ibm.com/developerworks/library/ba-data-mining-techniques

[39] Geoffrey E. Hinton, Learning multiple layers of representation, Trends Cogn. Sci. 11 (2007) 428–434

[40] Vikas C. Raykar, Ramani Duraiswami, Balaji Krishnapuram, A fast algorithm for learning a ranking function from large-scale data sets, IEEE Trans. Pattern Anal. Mach. Intell. 30 (7) (2008) 1158–1170, 200.

[41] Myra Spiliopoulou, Michael Hatzopoulos, Yannis Cotronis, Parallel optimization of large join queries with set operators and aggregates in a parallel environment supporting pipeline, IEEE Trans. Knowl. Data Eng. 8 (3) (1996) 429–445.

[42] Nicholas D. Lane, Ye Xu, Hong Lu, Andrew T. Campbell, Tanzeem Choudhury, Shane B. Eisenman, Exploiting social networks for large-scale human behavior modeling, IEEE Pervasive Comput. 10 (4) (2011) 45–53

[43] Björn Bringmann, Michele Berlingerio, Francesco Bonchi, Aristides Gionis, Learning and predicting the evolution of social networks, IEEE Intell. Syst. 25 (4) (2010) 26–35.

[44] Hao Ma, Irwin King, Michael Rung-Tsong Lyu, Mining web graphs for recommendations, IEEE Trans. Knowl. Data Eng. 24 (12) (2012) 1051–1064.

[45] Daniel A. Keim, Christian Panse, Mike Sips, Visual data mining in large geospatial point sets, IEEE Comput. Graph. Appl. 24 (5) (2004) 36–44.

[46] David Thompson, Joshua A. Levine, Janine C. Bennett, Peer-Timo Bremer, Attila Gyulassy, Valerio Pascucci, Philippe P. Pébay, Analysis of large-scale scalar data

using hixels, in: 2011 IEEE Symposium on Large Data Analysis and Visualization (LDAV), 2011, pp. 23–30.

[47] SIlya Molchanov, Theory of Random Sets, Springer, 2005

[48] Shun-Feng Su, Chen-Chia Chuang, C.W. Tao, Jin-Tsong Jeng, Chih-Ching Hsiao, Radial basis function networks with linear interval regression weights for symbolic interval data, IEEE Trans. Syst. Man Cyber.–Part B: Cyber. 19 (6) (2011) 1141–1151

[49] https://aws.amazon.com/about-aws/whats-new/2006/08/24/announcing-amazon-elastic-compute-cloud-amazon-ec2---beta

[50] Steve Loughran, Jose Alcaraz Calero, Andrew Farrell, Johannes Kirschnick, Julio Guijarro, Dynamic cloud deployment of a mapreduce architecture, IEEE Internet Comput. 16 (6) (2012) 40–50.

[51] https://en.wikipedia.org/wiki/Cloud_computing

[52] https://en.wikipedia.org/wiki/Bio-inspired_computing

[53] http://www.cse.wustl.edu/~jain/cse574-16/ftp/biocomm/index.html

[54] Gershenfeld, Neil; Chuang, Isaac L. (June 1998). "Quantum Computing with Molecules" (PDF). Scientific American.

[55] https://www.datapine.com/blog/big-data-examples-in-healthcare/

[56] https://www.talascend.com/industry-focus/big-data-and-information-technology

[57] https://en.wikipedia.org/wiki/Big_data#International_development

[58] Henry D. Morris Simon Ellis Jill Feblowitz Kimberly Knickle Marcus Torchia A Software Platform for Operational Technology Innovation" Retrieved 8 October 2017.

[59] Lee, Jay, Wu, F., Zhao, W. Ghaffari, M. Liao, L (January 2013). "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications". Mechanical Systems and Signal Processing

[60] Couldry, Nick, Turow, Joseph (2014). "Advertising, Big Data, and the Clearance of the Public Realm: Marketers' New Approaches to the Content Subsidy". International Journal of Communication. 8: 1710–1726.

BIOGRAPHIES

"Rupali Yadavrao Charde"
Master in Computer Science & Technology, Rajiv Gandhi Proudyogiki vishwavidyalaya, Bhopal, Madhya Pardesh, India (Pursuing)

"Triveni Yadaw"
Master in Computer Science & Technology, Rajiv Gandhi Proudyogiki vishwavidyalaya, Bhopal, Madhya Pardesh, India (Pursuing)

"Anand kumar"
Master in Computer Science & Technology, Rajiv Gandhi Proudyogiki vishwavidyalaya, Bhopal, Madhya Pardesh, India (Pursuing)

"Akash Sood"
[4]NPD Engineer, Technocrat Connectivity System, Delhi, India

"Sidharth Singh"
Bachlor of Computer Science Engineering, Maharishi Markandeshwar University, Mullana, Haryana, India(Pursuing)

"Riya Sahu"
SONT, Rajiv Gandhi Proudyogiki vishwavidyalaya, Bhopal, Madhya Pardesh, India