Neha Sharma et al. International Journal of Recent Research Aspects ISSN: 2349-7688,
Vol. 5, Issue 1, March 2018, pp. 127-132

Big Data Analytics in Health Care: A Literature Survey

Neha Sharma[*1], Arvind Panwar[#2], Urvashi Sugandh[%3]
[1,2]Department of Information Technology, Northern India Engineering College, New Delhi, India
[1]neha.sh.2689@gmail.com
[2]arvind.panwar@niecdelhi.ac.in
[3]HMRITM, New Delhi, India

Abstract - Healthcare industry data is increasing day by day, which compelled the industry to use big data analytic techniques for the analysis of the past data. Record keeping is important as whenever the patient comes, healthcare department need to track the previous records of the patient. According to the history of the patient, decision is made. Use of big data increases the productivity of the system. Working with big data is challenging but this is the need of the today's industry. Huge volume of data is created on daily basis, and this much amount of data is helpful in decision making process. Using big data in healthcare industry will decrease the manpower which in result will decrease the extra cost incurred in healthcare sector. Insight into the records, using big data helps stakeholders to get information regarding what can be done to improve personalized care, avoiding extra expenses, looking into the feedback of patient. This paper focuses on what are the applications of big data and how it is useful in healthcare industry. The paper also shows what are the merits and demerits of the same in healthcare industry.

Keywords - Hadoop, mapreduce, healthcare, prediction.

I. INTRODUCTION

Big data is on boom these days and as a result, this is the reason that it is the latest research topic. Big data is the need of future. In every industry, billions of records are generated on the daily basis. These records need to be analyzed, and higher-level management needs to make critical decisions based on them. Big data helps in decision making process in easy manner. It helps in handling the new challenges and prepares one to create the solutions for the problems which may arrive in near future. Big data analytics helps in discovering valuable decisions by understanding the data patterns and the relationship between them with the help of machine learning algorithms[1].

Big data has transformed the way of accepting things in the business, research and managerial activities. Data in the healthcare industry is huge and decision making on that is a difficult task. The healthcare industry allover the world focuses on reducing the expenses and efficiently handling the resources with improving patient care. Patient care is the major focus of healthcare industry but not even a single industry likes that it should to be done on the increasing cost. So, primarily focus is on improving patient care with less expense.

The data which will be needed to collect and analyzed by healthcare organizations may come from hospitals, ambulatory care facilities, wellness centers, referral networks, labs and imaging centers, research and other nontraditional data sources [2]. Doug Laney used volume, velocity and variety, known as 3Vs [3], to characterize the concept of Big Data. The term volume is the size of the data set, velocity indicates the speed of data in and out, and variety describes the range of data types and sources.

There are number of things, one need to keep track in the healthcare system. Doctor writes prescription, there are details of mediclaim, laboratory tests and everything need to be recorded for further enquires. Emergency care information need to be updated as it might be required anytime. The record kept in the big data is analyzed by examining large amount of data. Data from different operational systems is collected, and then data is transformed and loaded into the big data pool. Then by applying different techniques such as artificial intelligence, data mining, natural language processing, and predictive analytics, decision making is performed. Using these techniques, inherent patterns, anomalies etc can be discovered.

This paper focuses on big data analytics applications and how big data is helpful in healthcare industry. The paper is divided into different sections, section II explains about big data, where as section III focuses on 3 Vs of big data. Section IV explains research methodology. Section V provides the insight of distribution of articles by year of publication. Section VI explains the use of big data in healthcare. Section VII explains big data architecture and section VIII provides the conclusion and future directions.

II. WHAT IS BIG DATA?

Big data is angrowing term who describes massive amount of unstructured, semi-structured and structured data that has the potential to be mined for information. Traditional data processing software are not capable of dealing

with big data. The major challenges of big data are capturing data,storing data and analyzing the same, then sharing, transferring, querying,updating timely.

Big data consists of different types of data such as machine generated automatic data, social websites data, organizational data, ERP, sensors data etc.

With expanded data provisions, such as scientific experiments, sensor networks, telescopes, and high throughput instruments, the datasets growth at exponential rate [4,5]. The off-the-shelf techniques and technologies that we ready used to store and analyze data cannot work efficiently and satisfactorily. The challenges arise from data capture and data duration to data analysis and data visualization. In many instances, science is lagging the real world in the capabilities of discovering the valuable knowledge from massive volume of data. Basedon precious knowledge, we need to develop and create new techniques and technologies to excavate Big Data and benefit ourspecified purposes.

Big Data has reformed the way that we accept in doing businesses, managements and researches. Data-intensive science especially in data-intensive computing is coming into the world that aims to provide the tools that we need to handle the BigData problems. Data-intensive science [6] is evolving as the fourth scientific architype in terms of the previous three,namely empirical science, theoretical science and computational science. Thousand years ago, scientists describing the natural phenomenon only based on human empirical evidences, so we call the science at that time as empirical science. It is alsothe beginning of science and classified as the first paradigm. Then, theoretical science emerged hundreds of years ago as thesecond paradigm, such as Newton's Motion Laws and Kepler's Laws. However, in terms of many complex conceptand problems, scientists have to turn to scientific replications, since theoretical analysis is highly complicated and sometimes unavailable and infeasible.Afterwards, the third science paradigm was born as computational branch. Simulationsin large of fields generate a huge volume of data from the experimental science, at the same time, more and more large datasets are generated in many pipelines[7].

A. Applications of big data in different sectors:

1)Big Data in commerce and business

According to estimates, the volume of business data worldwide, across almost companies, doubles every 1.2 years [8].Taking retail industry as an example, we try to give a brief demonstration for the functionalities of Big Data in commercialactivities. There are around 267 million transactions per day in Wal-Mart's 6000 stores worldwide. For seeking for highercompetitiveness in retail, Wal-Mart recently collaborated with Hewlett Packard to establish a data warehouse which hasa capability to store 4 petabytes (see the size of data unit in Appendix A) of data, i.e., 4000 trillion bytes, tracing every purchase record from their point-of-sale terminals. Taking advantage of classy machine learning techniques to discoverthe knowledge hidden in this huge volume of data, they successfully improve efficiency of their pricing strategies and advertising campaigns. The management of their inventory and supply chains also significantly benefits from the large-scalewarehouse.

2)Big Data in society administration

Public administration also involves Big Data problems [9]. On one side, the population of one country usually is verylarge. For another, people in each age level need different public services. For examples, kids and teenagers need more education; the elders require higher level of health care.

3)Big Data in scientific research

Many scientific fields have already become highly data-driven [10,11] with the development of computer sciences. Forinstance, astronomy, meteorology, social computing [12], bioinformatics [13] and computational biology [10] are greatlybased on data-intensive scientific discovery as large volume of data with various types generated or produced in these science fields [12]. How to probe knowledge from the data produced by large-scale scientific simulation? It is a certain Big Dataproblem which the answer is still unsatisfiable or unknown.

For instances, a sophisticated telescope is regarded as a very large digital camera which generate huge number of universal images. For example, the Large Synoptic Survey Telescope (LSST) will record 30 trillion bytes of image data in asingle day. The size of the data equals to two entire Sloan Digital Sky Surveys daily. Astronomers will utilize computingfacilities and advanced analysis methods to this data to investigate the origins of the universe. The Large Hadron Collider(LHC) is a particle accelerator that can generate 60 terabytes of data per day. The patterns in those data can give us anunprecedented understanding the nature of the universe. 32 petabytes of climate observations and simulations were conserved on the discovery supercomputing cluster in the NASA Center for Climate Simulation (NCCS). The volume of humangenome information is also so large that decoding them originally took a decade to process. Otherwise, a lot of other e-Science projects [10] are proposed or underway in a wide variety of other research fields, range from environmental science, oceanography and geology to biology and sociology.

B. 3 Vs of big data:

Big data is categorized basically in 3 Vs:

1) ***Volume:*** The quantity of generated and stored data. The size of the data determines the value and potential insight- and whether it can actually be considered big data or not.
2) ***Variety:***The type and nature of the data. Complexity, thousands or more features per data item, many data types, and many data formats refer variety of data.
3) ***Velocity:***In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development.

III. RESEARCH METHODOLOGY

Papers are collected from different sources Springer, IEEE and various national and international journals. First of all papers were searched for big data and then it was checked whether they belong to big data or not. If they didn't belong to big data, then the papers were discarded and searching continued. Figure 1 depicts the research methodology followed while writing this paper.

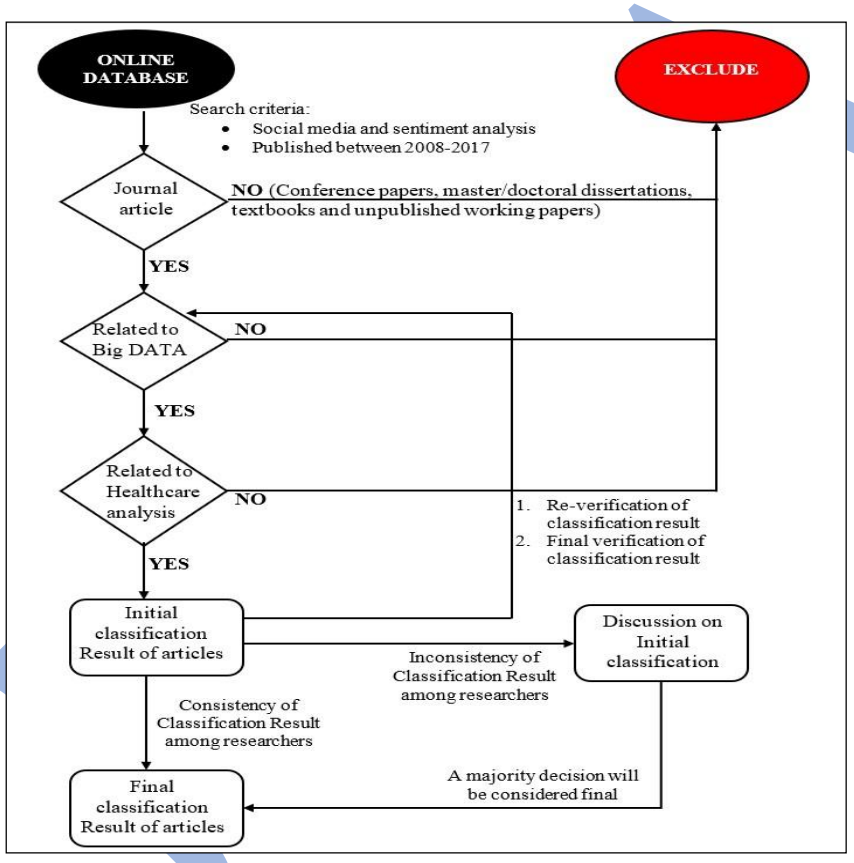

Fig. 1: Research Methodology

A. Distribution of articles by year of publication

Table 1: Distribution of Articles

S.NO.	Authors	Title of Paper	Literature Review	Year of Publication
1.	Richa Gupta, Sunny Gupta, Anuradha Singhal	Big Data : Overview"	This paper provides an overview on big data, its importance in our live and some technologies to handle big data. This paper also states how Big Data can be applied to self-organizing websites which can be extended to the field of advertising in companies [13].	2014
2.	Priya P. Sharma, Chandrakant P. Navdeti	Securing Big Data Hadoop: A Review of Security Issues, Threats and Solution	This paper discusses about the big data security at the environment level along with the probing of built in protections [12].	2014

3.	Wei Fan, Albert Bifet	Mining Big Data: Current St atus, and Forecast to the Future	The paper presents a broad overview of the topic big data mining, its current status,controversy, and forecast to the future. This paper also covers various interesting and state-of-the-art topics on Big Data mining [14].	2012
4.	Chanchal Yadav, Shullang Wang, Manoj Kumar	Algorithm and Approaches to handle large Data- A Survey	This paper presents a review of various algorithms from 1994-2013 necessary for handling big data set. It gives an overview of architecture and algorithms used in large data sets. These algorithms define various structures and methods implemented to handle Big Data and this paper lists various tools that were developed for analyzing them [15].	2013
5.	Puneet Singh Duggal, Sanchita Paul	Big Data Analysis : Challenges and Solutions"	This paper presents various methods for handling the problems of big data analysis through Map Reduce framework over Hadoop Distributed File System (HDFS). Map Reduce techniques have been studied in this paper which is implemented for Big Data analysis using HDFS [16].	2013
6.	Radu IoanCiobanu, Valentin Cristea, CiprianDobre And Florin	Big Data Platforms For The Internet Of Things	This paper focuses on how Big Data could change the research direction in the business model by providing services along with products. Technology shift generate more data through various applications like wireless sensors, smart devices, social mediaetc [17].	2014
7.	Flavio Bonomi, Rodolfo Milito, Preethi Natarajan And Jiang Zhu	Fog Computing: A Platform For Internet Of Things And Analytics	This paper proposed a hierarchical distributed architecture for IoT. Fog computing proposes a new breed of applications and services to have a productive interaction between existing cloud and Fog.Special focus given to Analytics and challenges of BigData. Fog computing is the next level of the cloud computing and it uses common resources. A Smart Traffic Light System(STLS) and Wind form systems were taken as the use cases in Fog computing [].	2014
8.	Mukherjee, A.; Datta, J.; Jorapur, R.; Singhvi, R.; Haloi, S.; Akram, W.	Shared Disk Big Data Analytics With Apache Hadoop	This paper discusses the necessity of Big Data and Big Data techniques which is required to process huge amount of data and to discover insights. Hadoop is a open source platform used for implementing Mapreducer Model. The performance of VERITAS Storage Foundation Cluster File System (SF CFS) is compared with Hadoop distributed file system (HDFS) for shared data Big Data analytics. Analytics with clustered file system is best suited for this proposed model [19].	2012
9.	Kudakwashe Zvarevashe1, Dr. A Vinaya Babu	Towards MapReduce Performance Optimization: A Look Into The Optimization Techniques In Apache Hadoop For Big Data Analytic	This paper provides the insight into the optimization techniques for Hadoop for big data analytics [20].	2014

IV. BIG DATA IN HEALTHCARE

Big data analytics represents a new approach to analytics. It does not yet have a large or significant footprint India or internationally [21]. However, the continuing digitization of health records together with the interoperable electronic health record (EHR), presents new opportunities to investigate a myriad of clinical and administrative questions. There is possible to layer BDA-type applications, in a privacy-protective style, on top of the initial health IT set-up to derive value that might not otherwise be found. What follows are some innovative ideas and solutions.

- Clinical decision support – BDA technologies that sift through large amounts of data, understand, categorize and learn from it, and then predict outcomes or recommend alternative treatments to clinicians and patients at the point of care[22].
- Personalized care – Predictive data mining or analytic solutions that can leverage personalized care (e.g., genomic DNA sequence for cancer care) in real time to highlight best practice treatments to patients. These solutions may offer early detection and diagnosis before a patient develops disease symptoms.
- Public and population health – BDA solutions that can minesocial media andweb-based data to guess flu eruptions based on consumers' search, social content and query activity. BDA solutions can also provisionepidemiologistsand cliniciansperforming analyses across patient populations and care sites to help recognize disease trends.
- Clinical operations – BDA can support initiatives such as wait-time management, where it can mine large amounts of historical and unstructured data, look for patterns and model various scenarios to predict events that may affect wait times before they actually happen.

- Policy, financial and administrative – BDA can support decision makers by integrating and analyzing data related to key performance indicators.

A. Benefits of Big Data Analytics in Healthcare Industry

Big Data analytics in healthcare can be used to increase the values in following fields:

- Public Health: By using the big data analytics, experts can analyze disease patterns and record disease outbreaks, public health issues can be improved with analytics approach.
- EMR(Electronic Medical Record): An EMR holds the standard (structured and unstructured) medical data that can be assessed with the big data analytic approach to guess patients at risk and deliver him effective care.
- Patient Profile Analytics: by applied Advanced analyticsto patient'sprofile for identifying individuals who could benefit from proactive approach.
- Genomic Analytics: The data analytic tactic can be efficiently included in genomic analytics to make this method a part of regular medical care decision process.
- Fraud Analysis: This data analytics approach helps analyze greater number of claim requests to curtail down fraud cases. An effective analysis can help reduce fraud, waste and abuse.
- Safety Monitoring: Data analytics can also be used to investigate real time great volumes of brisk data in hospitals. The approach may help in the safety monitoring and negative event forecast.

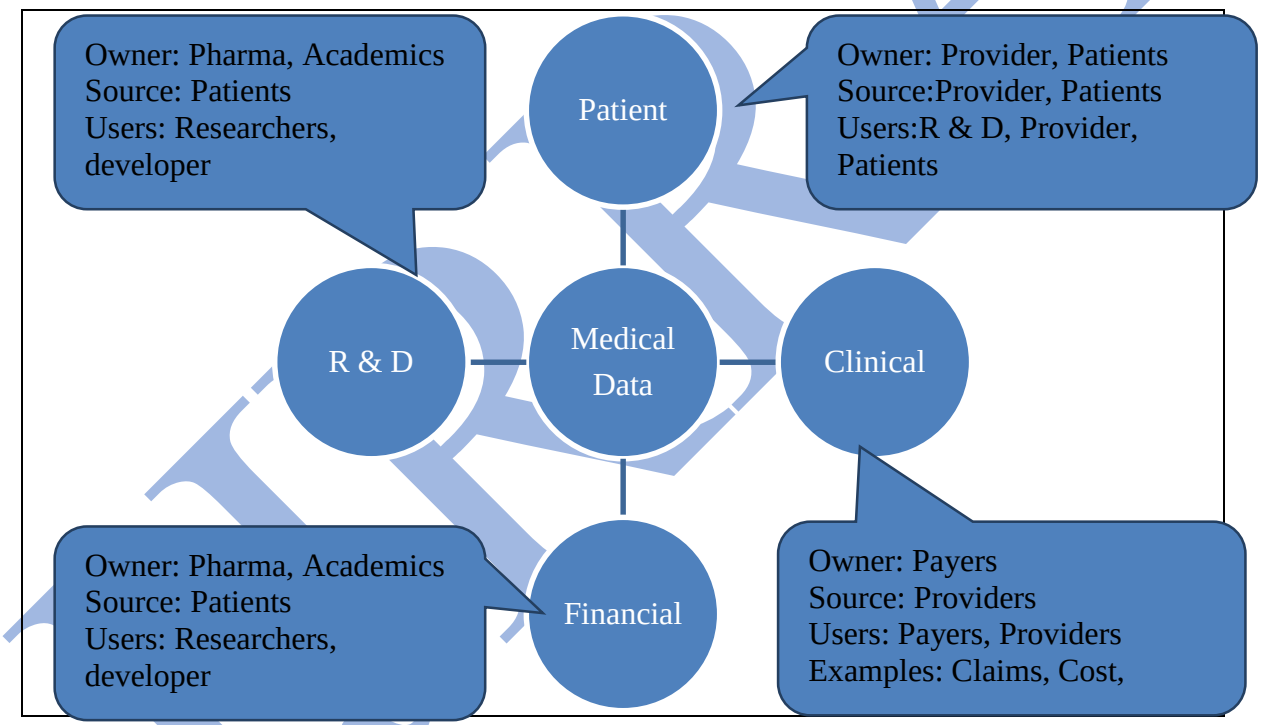

Fig. 2: type of medical data

B. Challenges of Big Data Analytics in Healthcare Industry

- Cleaning: In healthcare business every Healthcare providers are casuallywith the importance of cleanliness in the clinic and the operating room, but they may not be attentive of how important it is to cleanse their data, too. A murky data can quicklyspoil a big data analytics plan, particularly when data come from many different sources and format. Data cleaning guarantees that datasets are consistent, accurate, correct, relevant, and not corrupted in any way.
- Storage: Clinicians hardly think about where their data is being stored, but it's a critical cost, security, and performance issue for the IT department of healthcare service provider. According to time, data in healthcare grows exponentially, some healthcare providers are no longer able to manage the costs and effects data centers.
- Security: Providing the security for patient Data in healthcare is the number one priority for the organizations, especially when the world wake of a rapid-fire series of high profile breaches, hackings, and ransomware episodes. Healthcare data is issue to a nearly infinite array of vulnerabilities.
- Querying:Strong metadata and robust stewardship protocols also make it easier for organizations to query their data and get the expected answers. The capability to query data is foundational for reporting and

analytics, but healthcare companies must typically overcome a number of challenges before they can involve in meaningful analysis of their big data resources.

- Reporting: After providing good querying tool for data analytics in healthcare, next challenge is to provide a great report generation tool. Tool must generate a report that is clear, concise, and accessible to the target audience.
- Visualization: At the point of care, a clean and clear data visualization can make it much easier for a clinician to understand and interpret information and use it appropriately.
- Updating: Data in Healthcare is not static it changes time to time. Most elements in data will require frequent updates in order to retain current status of patient and relevant. For some datasets, like patient vital signs, these updates may occur every few seconds. Soits very important to update data on time.
- Sharing: In healthcare industry there are a numbers of service provider who deliver services to patient, some providers operatein a vacuum, and fewer patients receive all their care at a single location.

Letsdiscuss what is medical data and type of medical data.Whenever any person goes to a doctor, a lot of data is stored, processed, collected, analyzed, or disseminated. A big amount of that data may contain numbers related to your health, such as your heart rate during a particular visit. But there is potentially more to it than that.
Generally speaking, medical (clinical) data refers to health-related information that is related with systematic patient care or as part of a clinical trial program. There are many categories of such data, as this is a pretty broad definition. Figure 2 shows different type of medical data.

V. Conclusion

The paper shows the how big data is useful in health care industry and how the need of big data is increasing day by day. Big data helped in resolving number of things in an efficient way to improve the healthcare industry. Things like as Predictive readmissions, analyzing test variances, rapid bedside response, tracking patients waiting time, home health monitoring, chronic disease management and patient's scorecard generation has become the easy thing to calculate. Calculation of these things helps in analysis of healthcare system and further reduces the decision-making burden as OLAP, Data Mining like tools can be directly applied on them. But if he healthcare system is too small for some organization, then use of big data increases the complexity. So, in that case it is better to use conventional methods, instead of using big data.

References

[1] Priyanka K., Prof NagarathnaKulennavar, "A Survey On Big Data Analytics In Health Care ",IJCSIT, Volume 5(4), ISSN 0975-9646.
[2] " Data-driven healthcare organizations use big data analytics for big gains", IBM White Paper.
[3] Doug Laney, 3d Data management: controlling data volume, velocity and variety, Appl. Delivery Strategies Meta Group (949) (2001).
[4] Clifford Lynch, Big data: how do your data grow?, Nature 455 (7209) (2008) 28–29
[5] Alex Szalay, Jim Gray, Science in an exponential world, Nature 440 (2006) 23–24
[6] Gordon Bell, Tony Hey, Alex Szalay, Beyond the data deluge, Science 323 (5919) (2009) 1297–1298.
[7] Tony Hey, Stewart Tansley, Kristin Tolle, "The fourth paradigm: data-intensive scientific discovery", Microsoft Research (2009).
[8] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, "Big data: The Next Frontier forInnovation, Competition, and Productivity", McKinsey Global Institute, 2012.
[9] Randal E. Bryant, Data Intensive supercomputing: The Case for Disc. Technical Report CMU-CS-07-128, 2007.
[10] Randal E. Bryant, Data-intensive scalable computing for scientific applications, Comput. Sci. Eng. 13 (6) (2011) 25–33
[11] Alexander S. Szalay, Extreme data-intensive scientific computing, Comput. Sci. Eng. 13 (6) (2011) 34–41
[12] Priya P. Sharma, Chandrakant P. Navdeti, (2014), " Securing Big Data Hadoop: A Review of Security Issues, Threats and Solution", IJCSIT, 5(2), pp2126-2131
[13] Richa Gupta, Sunny Gupta, Anuradha Singhal, (2014), "Big Data:Overview", IJCTT, 9 (5)
[14] Wei Fan, Albert Bifet, "Mining Big Data: Current Status and Forecast to the Future", SIGKDDExplorations, 14 (2), pp1-5
[15] Chanchal Yadav, Shullang Wang, Manoj Kumar, (2013) "Algorithm and Approaches to handle large Data- A Survey", IJCSN, 2(3), ISSN:2277-5420(online), pp2277-5420
[16] Puneet Singh Duggal, Sanchita Paul, "Big Data Analysis:Challenges and Solutions", Int. Conf. on Cloud, Big Data and Trust, RGPV,2013.
[17]. Radu -Ioan,Ciobanu,ValentinCristea, CiprianDobre and Florin Pop,Big Data Platforms for the Internet of Things,2014,Springer
[18] Flavio Bonomi, Rodolfo Milito, Preethi Natarajan and Jiang Zhu,Fog Computing: A Platform for Internet of Things and Analytics,springer(2014)
[19] Mukherjee, A.; Datta, J.; Jorapur, R.; Singhvi, R.; Haloi, S.; Akram, W. "Shared disk big data analytics with Apache Hadoop"(18-22 Dec. 20012)
[20] Kudakwashe Zvarevashe1, Dr. A Vinaya Babu,Towards MapReduce Performance Optimization: A Look into the Optimization Techniques in Apache HadoopforBigData Analytics(2014)
[21] W. Raghupathi and V. Raghupathi, ``Big data analytics in healthcare:Promise and potential,"Health Inf. Sci. Syst., vol. 2, no. 1, pp. 1#10, 2014.
[22] O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis, and K. Taha,``Ef#cient machine learning for big data: A review,"Big Data Res., vol. 2,no. 3, pp. 87#93, Sep. 2015.

