

## Applications of Big Data in the Digital India: Opportunities and Challenges

<sup>1</sup> **Vinay Kumar**

Professor, Vivekananda School of IT, VIPS,  
GGSIPU, New Delhi 110 088, India.

<sup>2</sup> **Arpana Chaturvedi**

Asst.Professor, Jagannath International Management School,  
JIMS, GGSIPU, New Delhi 110070,India.

<sup>3</sup> **Poonam Verma**

Asst.Professor, Jagannath International Management School,  
JIMS, GGSIPU, New Delhi 110070,India.

---

DOI: <http://dx.doi.org/10.21013/jte.v3.n3.p7>

**How to cite this paper:**

**Kumar, V., & Chaturvedi, A.** (2016). Applications of Big Data in the Digital India: Opportunities and Challenges. *IRA-International Journal of Technology & Engineering* (ISSN 2455-4480), 3(3). doi:<http://dx.doi.org/10.21013/jte.v3.n3.p7>

---

© Institute of Research Advances



This work is licensed under a [Creative Commons Attribution-Non Commercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/) subject to proper citation to the publication source of the work.

**Disclaimer:** The scholarly papers as reviewed and published by the Institute of Research Advances (IRA) are the views and opinions of their respective authors and are not the views or opinions of the IRA. The IRA disclaims of any harm or loss caused due to the published content to any party.

---

**ABSTRACT**

*With the advent of Social Networking Sites (SNS), volumes of data are generated daily. Most of these data are multimedia type and unstructured with exponential growth. This exponential growth of variety, volume and complexity of structured and unstructured data leads to the concept of big data. Managing big data and harnessing its benefits is a real challenge. With increase in access to big data repository for various applications, security and access control is another aspect that needs to be considered while managing big data. We have discussed area of application of big data, opportunities it provides and challenges that we face in the managing such huge amount of data for various applications. Issues related to security against different threat perception of big data are also discussed.*

**Keywords:** Big Data, HDFS, IDC, Sqoop, Hive, Pig, Digital India, Governance, GFS, Hadoop, SME, Map Reduce, Analytics

**1. Introduction**

Big data implies large amount of data collected and stored in either structured or unstructured form or in both. The types of data may include business transactions, e-mail messages, photos, surveillance videos, activity logs and unstructured text from blogs and social media, medical, banks, e-Governance services, media houses, defense, corporate sectors as well as the huge amounts of data that can be collected from various sensors like CCTV. If this data effectively cleaned and analyzed, can help in data visualization of trends in business for various enterprises.[1] This emerging technology has the potential of influencing all walks of digital life. Data from tweets, blogs, product evaluations and other social networks sites facilitates an entrepreneur to analyze consumer’s preferences and products information. It opens a wide range of possibilities for *organizations to understand the needs of their customers predict their preferences and demands.* [2][12]

‘Digital India’ program is introduced in India with the vision to transform India into a digitally empowered society and knowledge economy. Its vision is broadly centered around three key areas: i) Infrastructure as a utility to every citizen, ii) Governance and services on demand and iii) Digital empowerment of citizens. Consequently availability of various Data resources has increased. These data resources can provide an opportunity to the business enterprises to explore the business trends and turn investments into huge profitable returns.

We discuss various applications of big data in digital India and what opportunities are available to the different stake holders. We also discuss challenges that are being faced and those challenges which are likely to be faced. The big data is characterized broadly by five V’s attributes: Volume, Velocity, Variety, Veracity and Values.[1]

*Volume* implies the amount of data it contains. It grows exponentially over a period of time and can reach to anything from Exabyte to Bronto bytes and even beyond. If we take all the data generated in the world up to 2015, the same amount of data is likely to be generated every minute in the coming days. This large data poses challenge to store and analyse using traditional database technology. New big data tools use distributed systems so that we can store and analyse data across databases that are geographically scattered around and linked with network.

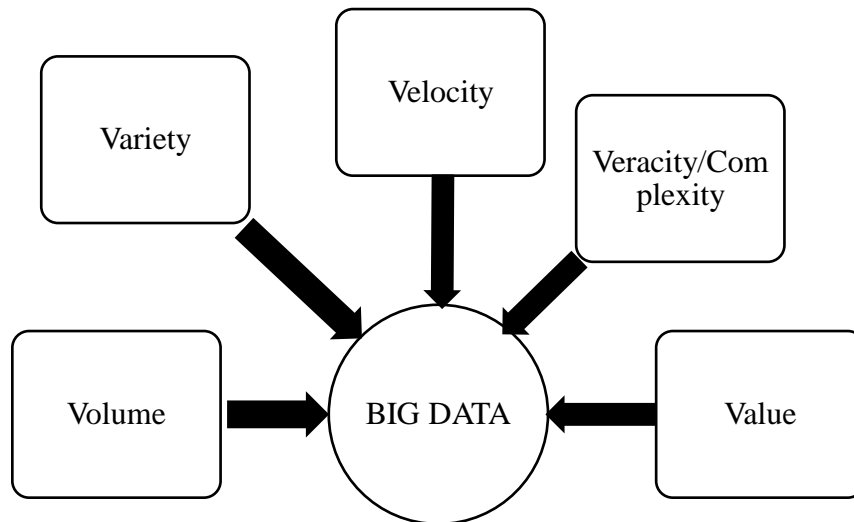


Figure 1: Characteristics of Big Data

*Velocity* in this context means that data moves through various sources such as online systems, sensors, social media, web click stream capture, and other channels. Think of social media messages going viral in seconds. Such scenario of big data refers to its velocity. Sometimes it needs to be analysed on the fly without storing it. [1]

Big data consists of *Variety* of types of data from many sources. Data are structured, semi-structured, as well as unstructured (think emails, text messages, etc.). Structured data neatly fits tables and traditional tools are capable to deal with it. With the advent of Social Media (SM), online sensors, Remote Sensing satellites, 80% of the data generated these days are unstructured (text, images, video, voice, etc.). Big data technology facilitates integration of different types of data such as messages, social media conversations, photos, sensor data, and video or voice recordings.

*Veracity also called Complexity* implies that the data must be able to traverse multiple data centers, the cloud and geographical zones. Handling, managing, analyzing such tremendous amount of data is very complex. Its semi structured, quasi structured nature increases the complexity in managing, storing as well as in accessing.

*Value* is measure of tangible or intangible benefits an organisation may gain from Big Data. Having access to big data is no good unless we can turn it into value. Companies are generating amazing value from their big data. A schematic diagram in shown is the Figure 1, depicting relationship of big data and its five characteristics. Due to these five characteristics of big data, it has potential for its applicability in many areas of application. To mitigate challenges, a comprehensive strategy and program plan for big data integration and better applications are required. The wide variety of NoSQL tools, developers and the current status of promote and market are creating uncertainty within the data management background. It consequently poses challenges. The paper is organized in five sections. Section 2 deals with literature review. Its area of application is discussed in Section 3. The challenges are outlined in Section 4 and paper is finally concluded in the Section 5.

## 2. Literature Review

Close to 2.5 quintillion bytes of data is generated daily. Based on this statistical figure, IDC report on Digital Universe predicts that digital universe will approximately double every 2 years. IDC has further predicted that big data will grow @ 27% CAGR (Compound Annual Growth Rate). In monetary terms it comes out to be of \$41.5 billion through 2018, which is about 6 times the growth rate of the overall information and communication technology market.[12] Wikibson is predicting

revenues of \$53.4 billion by 2017, as more business begin to realize real benefit from big data analytics. In fact by 2020 IDC believes that big data will be one of the key drivers for the economic growth of any nation worldwide. [24]According to IESA, the country's ESDM market will touch \$400 BN by 2020. By 2020, this figure is expected to cross 50 percent the way m-commerce is exploding.[25]

India alone has 173 million social media users and also world's second largest Face book audience, India will surpass 100 million mobile Face book users by 2018 as per e-Marketer report. It is also projected that 90% of internet users also use Face book for ease of accessibility. People from all over the world upload up to 50 million pages on Face book every day. Nearly 40% of the internet users also have to access to the internet from their smart phones that helps people to get 24/7 access to information. More than 1.2 billion broadband connections are there as per a survey and 80 percent of them do not have net banking access. Approx 20-30 percent of the global population can be termed as data users.[2][12]

GE worked with Italian Airlines – Alitalia, to monitor wing-flap positions and relay adjustments during landing and fuel usage, leading to an estimated savings of USD 46 million. Their ultimate goal is to give airlines and pilots real time suggestions to improve plane efficiency through changing trajectory or velocity during flight and identifying early signs of faulty parts, improving safety, reliability and cost. In another study, Hitachi converted their Shinkansen Bullet train control system that has been running since the 1964 Tokyo Olympics without any fatalities to the European Train Control systems. By using Big data Analytics of the sensors built into the train and the train control system.[2][12]

According to recent study by NASSCOM and BLUEOCEAN, the analytics market in India is expected double from \$1 billion in last fiscal to \$2.3 billion by the end of 2017-18. According to Analyst Firm McKinsey and Co., by 2018, the United States alone could face a shortage of 140,000 to 19,000 people with deep analytical skills. 1.5 Million Managers and analyst with the know-how to use analysis of Big Data to make effective decisions. In 2012, Gartner analyst predicted that by 2015, 4.4 millions of IT Jobs globally will be created to support Big Data with 1.9 Millions of those jobs in United States. [16]

Big Data is real and is here to stay, no nation can afford to close its eyes and ignore this phenomenon. India, with its vast pool of IT talent and industries, is placed at right place, right time to join the bandwagon of 'The Big Data platform,' only the right intent is required. It has so many challenges and opportunities as well which is required to be handled. Google file system (GFS) is used to handle challenges of wide web scale storage providing client with familiar OS-level byte stream abstraction [15]. [17] The processing of such large scaled files can be easily processed using Map Reduce programming model and platform [13][15][17][23]. Apache Hadoop Map Reduce platform file system HDFS is now popularly used for use cases, web indexing, clickstreams and log analyses. Hadoop developed a high declarative language like Pig for Yahoo![22],hive from Face book[18],Jaql from IBM[28].Pig is similar to relational algebra and is reportedly used for over 60%of Yahoos! Map Reduce use cases, Hive is SQL like and reported to be used for over 90% of the Face book Map Reduce use cases. Microsoft has also announced that its future "Big Data" strategy includes support for Hadoop[20][15].Microsoft technologies uses parallel runtime systems called Dryad and two higher level programming models Dryad LINQ and SQL like SCOPE language[21] which utilizes Dryad as well[15].

### 3. Applications of Big Data

Big data analytics is the process of examining large data sets containing a variety of data types, what we refer here as big data, to uncover hidden patterns, unknown

correlations, market trends, customer preferences and other useful business information. Every aspect of our lives will be affected by big data. However there are some areas where big data is already making a real difference today and it is adding real value in it. Big data creates social impact by enriching the human lives. Areas like transportation, energy, public safety, urban development, soil fertility analysis, analysis of water retention strength of soil etc., are finding the concept of big data analytics more helpful. It unearths unexplored characteristics of the data and helps in taking decision 'out of the box' sometimes. The big data is very useful in education, training and educational research wherein repository of large data in various forms are fetched and organized for presentation purposes. Defense is another area where big data has tremendous scope. Some application areas are briefly mentioned below.[4]

*Predictive Model for understanding end users behaviour and preference:* Understanding customer's behaviors and preferences is important to expand the market share of any product of a company. Social media sites are rich source for such data. Analysis of such big data gives a complete picture of customers preferences, behaviors etc. A predictive model is used to implement it. Study shows that U.S. retailer Target, accurately predict using the big data analytics that when one of their customers will expect a baby. A telecom company may predict customer churn, Wal-Mart may predict what products will sell and so on. The concept of big data analytics is used to improve e-Democracy by reaching to the people according their preferences. It is believed that Obama's win in the 2012 presidential election campaign was due to his team's superior ability to use big data analytics. [2][12]

*Optimization models for different Business processes:* Various Business processes can be optimized using better predictive models which provide effective optimization tools. Using the predictive model on the data from social media data, web search trends and weather forecasts, retailers may optimize their stock. Few Big data tools improve HR Business process as can measure an organisation's culture and its manpower engagements. Big data analytics are extensively used for Business processes like supply chain or delivery route optimization. Many geographic positioning and radio frequency identification sensors are used to track goods or delivery vehicles and optimize routes by integrating live traffic data.

*Benefitting Individual quantification and performance optimization:* Individuals can be benefited by the data generated from wearable devices such as smart watches or smart bracelets. The armband collects data on wearer's calorie consumption, activity levels, and sleep patterns. While it gives individuals rich insights into health condition, the real value is in analyzing the collective data for better monitoring and diagnostic purposes. Analyzing such volumes of data will bring entirely new insights and may work as feedback to the individual.

*Analytics to provide better Healthcare and public health:* The computing power of big data analytics enables us to decode entire DNA strings in minutes. It helps in finding new cures and better understanding of a disease patterns. Big data technique is being used to monitor babies in a specialist premature and sick baby unit. By recording and analyzing every heart beat and breathing pattern of a baby, a baby care unit can predict infections 24 hours before appearance of any physical symptoms. Big data analytics allow us to monitor and predict the developments of epidemics and disease outbreaks. Integrating data from medical records with social media analytics enables us to monitor and control flu outbreaks. Wearable healthcare devices generate lots of data. If these data are stored and analyzed by a doctor, many problems related to health can be solved before they get out of control.

*Analytical tools to Improve sports performance:* The combination of more and better data, better analytical tools and better computing power has allowed the use of

analytics, predictive models to blossom in sports. Many of the big analytical tools have come in the sports world to monitor the performance of a player and provide feedback for improvements. Nowadays much better technology, improved advanced analytical methods and huge advances in reliable real-time video data capture like SportVU are used in improving sports overall performance. The IBM Slam Tracker tool for tennis tournaments, video analytics for football or baseball are some examples of such tools. Many elite sports teams also track athletes outside of the sporting environment using smart technology to track nutrition and sleep, as well as social media conversations to monitor emotional wellbeing.

*New possibilities to Improve science and research:* CERN, the Swiss nuclear physics lab with its Large Hadron Collider, the world’s largest and most powerful particle accelerator is the very good example of big data application. CERN experiments to unlock the secrets of our universe i.e., how it started and how it works, generate huge amounts of data of the size of 30 Petabytes. Science and research is currently being transformed by the new possibilities that big data provides.

*Optimization tools to improve Machine and device performance:* Webster Research Center Of Xerox Corporation is working to build microchips to optimize the processes used to fabricate new materials and devices. It will be able to improve process control, lower process temperature and achieve more abrupt multilayer material interfaces for improved electrical performance. Livermore researchers have worked to improve the accuracy, speed, throughput and reliability of computer controlled machine tools.[13] In fact nowadays big data tools are used to optimization of energy grids can also be done by collecting data from smart meters and using big data tools to perform analytics. The main idea or concept behind is used to optimize the performance of computers and data warehouses. Big data analytics help in making machines and devices smarter and more autonomous. The big data tools are used to operate Google’s self-driving car. The Toyota Prius is fitted with cameras, GPS as well as powerful computers and sensors to safely drive on the road without any human intervention

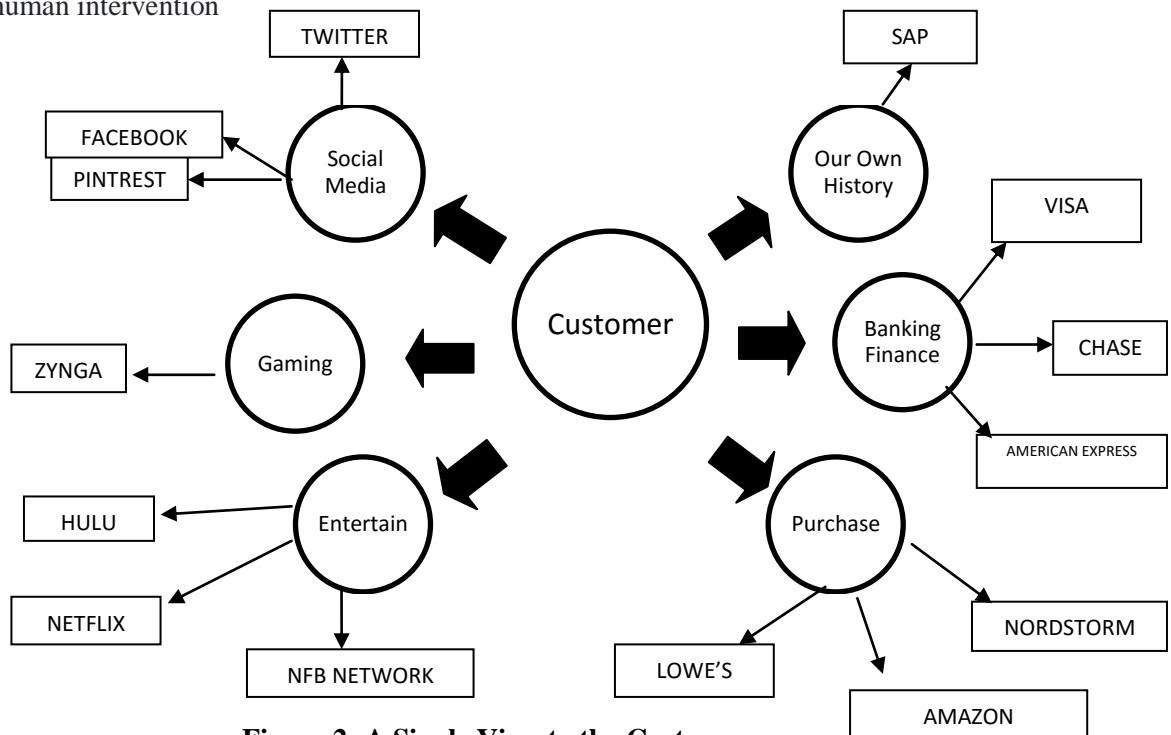


Figure 2: A Single View to the Customer

*Improving security and law enforcement:* Big data analytics security agencies in maintaining law and order from both internal and external threats. Recently National Security Agency (NSA) of USA used big data analytics to foil terrorist plots. The other examples application of big data in controlling crimes may include: detection and prevention cyber attack, controlling criminal activities by analyzing existing records and patterns, detection and avoidance of fraudulent transaction using credit card.

*Urban utility and transportation improvement:* Big data analytics is used to improve traffic flows in a city by analyzing integrated data about real time traffic information, social media and weather data. By integrating transport infrastructure and utility processes, number of cities is turned into Smart Cities, wherein a bus would wait for a delayed train and where traffic signals predict traffic volumes and operate to minimize jams. A striking example of Big Data adaption in the SME segment is the taxi-aggregators. The same in the case with small time travel operators and agents lying up with online portals which help in devising appropriate travel packages for the customers.

*To increase Financial Trading:* Trading organizations uses nowadays many analytical tools and employs many forms of analysis to help drive their trading strategies to optimize the value of their assets or speculate on future commodity price movements. Infact High-Frequency Trading (HFT) is an area where big data finds a lot of use today. Here, big data analytics are used to make decision related to majority of equity trading. It analyses trend from social media networks and news websites to take decision to buy and sell in real time.

#### 4. Challenges

With the advent of big data analytics and its utility in various walks of human lives, the task of managing this huge amount of data for the purpose of analysis, retrieving data and securing it from adversaries are posing big challenges in the way of big data applications. Although big data generates new growth opportunities for an industry, there are various challenges in big data that can be broadly viewed through two different approaches: Challenges in the various phases of processing of Big Data and Challenges in the different characteristics of Big Data. Phase cycle is depicted in the Figure 3. [26]

- *Challenges in the phase of data acquisition and recording* are ‘filtering of data’, ‘generating meta data’ and ‘data provenance analysis’. Filtered data should be intelligently processed and compressed to a convenient size. Right meta data must be automatically generated to describe the data which is recorded and measured. Regarding data provenance analysis, it is required to detect metadata dependencies for security and confidentiality application. This is computationally intensive and complex.
- *Challenges in the phase of information extraction and cleaning* are face during extraction of the required information from the databases and then present in the format required at a particular circumstance because in most case extraction is application dependent.
- *Challenges in the phase of data integration, aggregation, and representation* are associated with database design which depends on the outcome of the data analysis, data structures and semantics of data to be expressed in a form that is robust and host environment compatible. It becomes necessary to create effective database designs, either through devising tools for assistance by forgoing the design process completely and developing techniques to use databases effectively in the absence of intelligent database design.

- *Challenges in the phase of query processing, data modeling, and analysis* are faced due to lack of coordination between database systems, which host the data and provide SQL querying. Interconnected big data forms large heterogeneous information network. With such networks, information redundancy can be explored to compensate for missing data. Conflicting cases need to be crosschecked to validate trustworthy relationships. It also helps in uncovering hidden relationships and models. Data mining requires integrated, cleaned, trustworthy, and efficiently accessible data, declarative query and mining interfaces, scalable mining algorithms, and big-data computing environments.
- *Challenges in the phase of Interpretation* is encountered because of the expectation of rich visualization with assumptions and supplementary information. Interpretation involves examining all the assumptions made and retracing the analysis where there are possible sources of errors. Analytical pipelines involve multiple steps based on crucial assumptions behind the data recorded.

Out of the five characteristics of big data, we have outlined earlier, we face challenges due to three of them. These are volume, velocity, and variety. Security and privacy issues are magnified by velocity, volume, and variety of big data because of large-scale cloud infrastructures, diversity of data sources and formats, streaming nature of data acquisition, and high volume inter-cloud migration. Therefore, traditional security mechanisms, which are tailored to securing small-scale static data, as opposed to streaming data, are inadequate. Some major challenges in this area include secure computations in distributed programming frameworks; Secure data storage and transactions logs; End-point input validation and filtering; Real-time security compliance monitoring. Besides cryptographically enforced access control and secure communication, Granular access control, Granular audits and data provenance are other major challenges.

But at the same time this increased access to sensitive information organizations must take proactive measures to roll out a comprehensive and consistent identity and security management strategy. Today the physical world is seamlessly merging with the cyber space. Millions of people in the country rely on the services and information available in cyber space. Increasingly, the work of government, business and national infrastructure is being highly dependent on cyber space.

With the growing use of Information Technology in all walks of life, threats to cyber space have become one of the most serious economic and national security challenges for our country and the world at large. The cyber space being borderless and anonymous is exploited by malicious actors and adversaries for carrying out malicious activities in the cyber space. Cyber Security is also one of the key components envisaged for the success of the ambitious “Digital India” program of the government.



Figure 3: Challenges in various Phases of Processing of Big Data



Various challenges can impose serious risks in successful deployment of a big data program. Few risks can be inability to appropriately satisfy the growing community of data consumers, inability to actualize the program due to unmanageable data latency, flawed decision based on flawed data. Hence all the challenges can reflect different facets in the absence of a strategy for integrating big data into the enterprise environment.[26]

### **5. Recommendations for handling challenges:**

Firstly the big data talent gap has to be minimized and additional unknowns are to be introduced to get adopted universally the big data concept. The large scale and variety of data is to be absorbed and integrated in such a way that it is accessible globally. As the data is to be incorporated from diverse sources into an analytical platform, maintainence of the concurrency and consistency of data is to be handled properly. It is required to enable high-performance analytics so that used adequately by all different enterprises [19]. Advance analytics helps to find out possible solution for the various challenges and to discover more beneficial business opportunities, new business facts that no one in the enterprise knew before. The various technology and tools organizations can use are predictive analytics, discovery analytics, statistical analytics, machine learning, data visualization and natural language processing. Tool like Map Reduce over hadoop and HDFS can better help the organizations. Online learning algorithms can be used in Map Reduce. The model parameter learning in a learning algorithm can be viewed as shared global state which must be updated as the model is evaluated against training data. [5]

Streaming algorithm [16] can be used as an alternative model for dealing with large volumes of data with limited computational and storage resources. Learning algorithm based on linear classifier and multiclass classifiers can be used to design logistic regression model to make more accurate predict for various input parameters.[17] Multiclass classifiers based model can also be designed to do fraud detections in UIDAI scheme. The Map Reduce derives from providing an abstraction that allows developers to harness power of large clusters but abstraction manage complexity by hiding details and presenting required features to users.

The future directions suggested in the paper will be going to be dealt in the next paper. It will be very impactful and beneficial used as analytical tools and built on the existing paradigm of HDFS and Hadoop, overcoming the existing drawbacks in the current system.

### **6. Conclusion**

We have entered an era of Big Data. Through better analysis of the large volumes of data that are becoming available, there is the potential for making faster advances in many scientific disciplines and improving the profitability and success of many enterprises. However, many technical challenges described in this paper must be addressed before this potential can be realized fully. The challenges include not just the obvious issues of scale, but also heterogeneity, lack of structure, error-handling, privacy, timeliness, provenance, and visualization, at all stages of the analysis pipeline from data acquisition to result interpretation. These technical challenges are common across a large variety of application domains, and therefore not cost-effective to address in the context of one domain alone. Furthermore, these challenges will require transformative solutions, and will not be addressed naturally by the next generation of industrial products. We must support and encourage fundamental research towards addressing these technical challenges if we are to achieve the promised benefits of Big Data.

### Acknowledgment

We are very grateful to all those who have been constantly encouraging and providing required resources for such scientific research work besides the regular work which we are doing in our respective departments. We are grateful to our colleagues and seniors for their continuous support and encouragements and to all those who provided suggestions for improvement of this work.

### References

- [1] Posa Rambabu, 2015, August 14."Big data Introduction" Available: <http://www.journaldev.com/8734/introduction-to-bigdata>
- [2] Adeesh Sharma, "Making the right moves on Big Data", PCQUEST July 2015 pg. no. 37
- [3] Xueyuan Su, Garret Swart, Brian Goetz, Brian Oliver, Paul Sandoz, Oracle Corporation. Changing Engines in Midstream: A Java Stream Computational Model for Big Data Processing
- [4] NESSI White Paper, December 2012, Big Data, A new world of opportunities
- [5] H. Lee, K. J. Brown, A. K. Sujeeth, H. Cha\_, T. Rompf, M. Odersky, and K. Olukotun. Implementing domain-specific languages for heterogeneous parallel computing. IEEE Micro, 31(5):42{53, 2011
- [6] Efficient Capabilities of Processing of Big Data using Hadoop Map Reduce, Praveen Kumar1, Dr Vijay Singh Rathore, International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 6, June 2014.
- [7] Vinod Ganeshan,Head-BFSI Vertical, Hitachi Data Systems, " How Big Data is going to Impact Businesses in the Near Future" PC Quest, July 2015, Pg.No. 40-41
- [8] Aditya B. Patel, Manashvi Birla and Ushma Nair, "Addressing Big Data Problem Using Hadoop and Map Reduce," in Proc. 2012. Nirma University International Conference On Engineering, pp. 1-5.
- [9] "Big Data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute, May 2011, p. 11: [http://www.mckinsey.com/Insights/MGI/Research/Technology\\_andInnovation/Big\\_data\\_The\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/Insights/MGI/Research/Technology_andInnovation/Big_data_The_next_frontier_for_innovation).
- [10] Yogesh Dandawate,Senior Architect, Analytics and Allied Solution Cloud Move," Key steps in framing the Right Big Data Strategy" PC Quest, July 2015,pg. no. 38-39
- [11] Rashi Varshney "Big data: Analytics provides Key Opportunities for Growth to startups" PC Quest pg. No. 48-50
- [12] Karthik Sridhar, Founder & CEO data Culture ," Making Sense of Big data for SME's" PC Quest , July 2015, pg.no. 51
- [13] William A. Bookless "State of the Laboratory: Lawrence Livermore National Laboratory" Pg. No 78
- [14] Jeffrey Dean and Sanjay Ghemwat, Map Reduce: A Flexible Data Processing Tool, Communication Of the ACM,vol 53,Issue 1,jan 2010,pp 72-22
- [15] Puneet Singh Duggal and Sanchita Paul, Big Data Analysis: Challenges and Solutions, International Conference on Cloud Big Data and Trust 2013,Nov 13-15,
- [16] Carlos Ordonez, Algorithms and Optimizations for Big Data Analytics: Cubes, Tech Talks, University Of Houston, USA.

- [17] HADOOP-3759: Provide ability to run memory intensive jobs without affective other running tasks on the nodes. <https://issues.apache.org/jira/browse/HADOOP-3759>
- [18] PIGTutorial,yahooInc.,<http://developer.yahoo.com/hadoop/tutorial/pigtutorial.html>
- [19] [www.progress.com.big](http://www.progress.com/big) data challenges
- [20] Windows.Azure.Storage.<http://www.microsoft.com/windowsazure/features/storage/>
- [21] Dryad-Microsoft Research, <http://research.microsoft.com/enus/projects/dryad>
- [22] IBM-What is Jaql,[www.ibm.com/software/data/infosphere/hadoop/jaql](http://www.ibm.com/software/data/infosphere/hadoop/jaql)
- [23] 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
- [24] Onkar Sharma, “Demystifying Big Data” Dataquest, Feb 29,2016 pg. 19
- [25] Shrikanth G. ‘Make in India’ to empower the ESDN Industry, Data Quest Feb 29, 2016 pg. 22
- [26] David Loshin, President of Knowledge Integrity, Inc. “ADDRESSING FIVE EMERGING CHALLENGES OF BIG DATA”, White paper, <https://www.progress.com/docs/default-source/default-document-library/Progress/Documents/Papers/Addressing-Five-Emerging-Challenges-of-Big-Data.pdf>
- [26] “A community white paper developed by leading researchers across the United States”,<https://www.purdue.edu/discoverypark/cyber/assets/pdfs/BigDataWhitePaper.pdf>

**Authors’ Biography:**

**Vinay Kumar** is a Professor in Vivekananda Institute of Professional Studies, Delhi. Earlier he worked as Scientist in National Informatics Centre, MoCIT, and Government of India. He completed his Ph.D. in Computer Science from University of Delhi and MCA from Jawaharlal Nehru University, Delhi. He has authored a book on Discrete Mathematics and has contributed many research papers to refereed journals and conferences. His areas of interest are graph algorithm, information security & privacy and e-governance. He is member of CSI and ACM.

**Arpana Chaturvedi** is working as an Assistant Professor in Jagannath International Management School, Delhi. She is M.Sc. (Math), MCA and M. Phil. (Comp. Sc). She is pursuing PhD from Jagannath University. She is having 24 yrs of teaching and industrial experience in various organizations. She has contributed many research papers to referred journals and conferences. Her area of interest is programming, web technologies and database management.