International Journal of Advanced Research and Development
ISSN: 2455-4030, Impact Factor: RJIF 5.24
www.advancedjournal.com
Volume 2; Issue 3; May 2017; Page No. 143-146

Understanding the Issues and Concerns involved in big data

[1] Barkha Rani, [2] Kalpana, [3] Amit Kumar

[1] M.Tech, Computer Science, Delhi, India
[2] Research Scholar, (PhD), Center for Interdisciplinary Research in Basic Sciences, Jamia Milia Islamia, Delhi, India
[3] Assistant Professor, Department of Library and Information Science, Mizoram University, (A Central University) Aizawl, Mizoram, India

Abstract
Big Data has attracted the huge attention of professionals, researchers and scholars working in information Science field. Today, the world is flooded by information which has led the information industries to face the challenges like how to handle, organize, preserve for future and make accessible. To overcome these challenges, Big Data is emerging as a solution which is going to revolutionize information industry. The present paper is an effort to discuss the issues and concerns involved in Big Data.

Keywords: big data; data analytics; information industry; and information technology

1. Introduction
In today's era with the new dimensions taking place in each and every sector whether public or private, new ways of lifestyle, functioning and challenges have come up in picture for instance e- shopping, e-banking, e-governance, e-classes etc. Now, organizations are involved in data handling and its storage. To solve this issue Big Data has evolved as the solutions for handling the information with huge data. Therefore, Big Data has become one of the current and future research frontiers. We can be rightly say that Big Data will modernize many fields, such as business, information industry, scientific research, different public administration, and many more.

2. Big Data: What actually it is?
The Term Big Data in its current use was coined by Roger Magoulas [1], generally refers to data that exceeds the typical storage, processing, and computing capacity of conventional databases and data analysis techniques. As a resource, Big Data requires tools and methods that can be applied to analyze and extract patterns from large scale data. The rise of Big Data has been caused by increased data storage capabilities, increased computational processing power, and availability of increased volumes of data, which give organization more data than they have computing resources and technologies to process [2]. Oguntimilehin A, and Ademola E.O. define [3, 4] Big Data can be described as five Vs and a C. which are as follows:

a. **Volume:** Now we come to an age in which, it is quite complicated to speak about volume of the data in absolute sense. As technology increases, everything gets outdated; therefore it is good to think of volume in some related sense.
b. **Velocity:** data velocity refers to the speed of the creation of data, aggregation and streaming. E-commerce has grown rapidly speed as well as richness of the data which is used for various other business transactions. Management of data velocity is much more difficult problem than a bandwidth issues.
c. **Variety**: Big Data includes both structured and unstructured data. Now-a-days, It becomes a challenging task to manage, merge and for governing variant data.
d. **Variability**: quantity of data flows, in addition to the growing velocities and varieties in data, is highly variable.
e. **Value:** It is a subject to think ahead, that are we actually get good quality of data or is this data have any value for our business. And also, is after Big Data has came in existence our stirring troubles in industry has been get rectify or not? and
f. **Complexity:** As the data collected from many different sources, it becomes complex. However, it is compulsion to join and correlate relationships and make multiple data linkages.

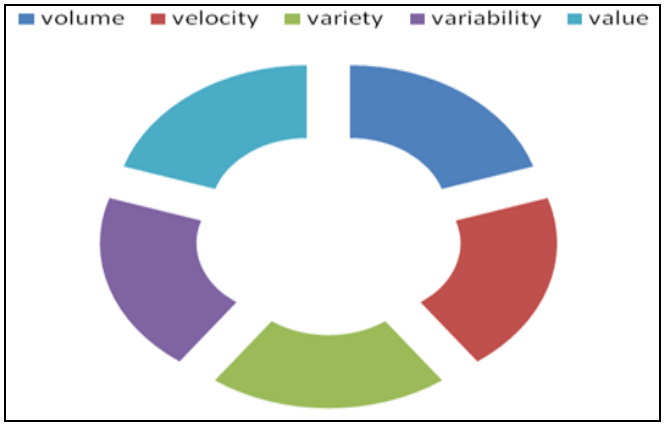

Fig 1: The five V's of big data

3. Big Data Process
Big Data process [5] involves many steps which play an important role in handling the information, are as follows

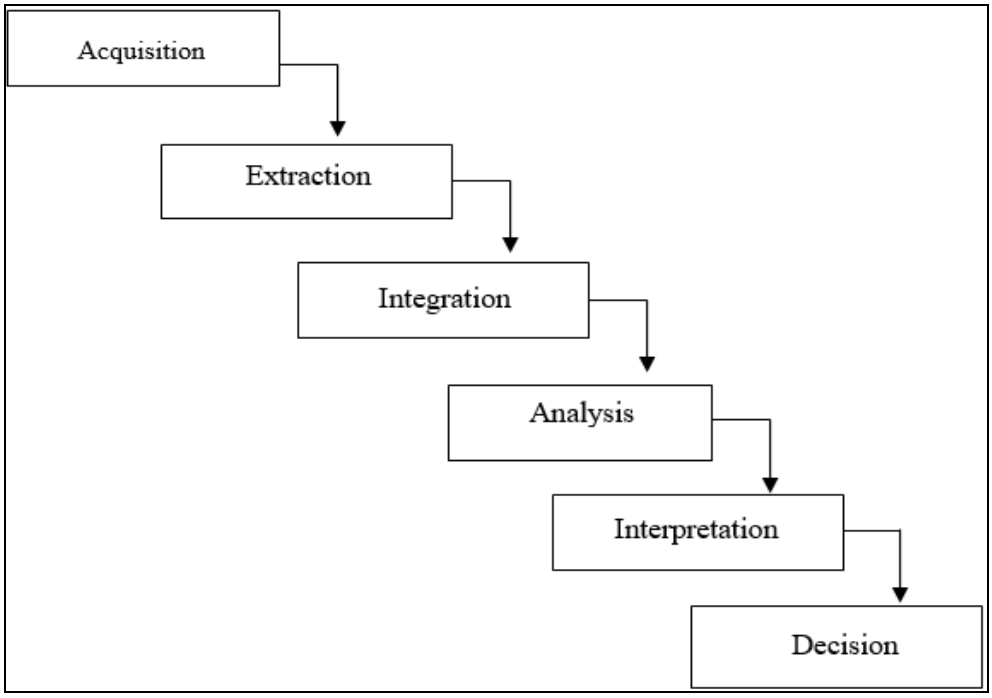

Fig 2

Table 1: Furthermore, the process of big data can be seen through the table given below

S. No.	Step	Process Include
1	Acquisition	Selection, Filtering, Metadata Generation And Managing Provenance.
2	Extraction	Transformation, Normalization, Cleaning, Aggregation, Error Handling.
3	Integration	Standardization, Conflict Management, Reconciliation, Mapping Definition
4	Analysis	Exploration, Data-Mining, Machine Learning, Visualization.
5	Interpretation	Knowledge of The Domain, Knowledge of The Provenance, Identification of Patterns of Interests, Flexibility of The Process.
6	Decision	Managerial Skills And Continuous Improvement of The Process.

4. Why Big Data is the need of the Hour

Today, the technology and architecture of Big Data have become an important aspect in the life of professionals working in information field and growing rapidly. Many important reasons behind the popularity of Big Data are there which makes it an important and indispensible aspect of 21st century. Some of the key enablers for the growth and popularity of Big Data are:

a. Advancement of ICT: As the advancements of ICT is growing day by day, the expectations and demands of the society are changing. Now the generation is techno savvy and they are habitual of handling the IT devices without any problem. So it is the responsibility of professionals working in information industry to make sure quality services and satisfy them with their changing demands.

b. Transformation of Information Industry and its Requirements: With the advancement of ICT the information industry is also changing. Now majority of the information is being produced in e-form with multimedia applications that has led the information industry to change their way of functioning and to adopt new technologies for to feed the information hunger of the society.

c. **Increase of storage capacities:** As information industry has completely changed and majority of the information is born digital now, now days have gone when the people were satisfied with the storage devices in GB (GigaBytes), because limited data and the format of the information born was not like as todays. As the increasing storage capabilities of users, information industries, entrepreneurs the need of the solution like Big Data is inevitable and it is helping to handle and organize the huge data/information. for example Google, Amazon etc.

d. Increase of processing power: Today, very sophisticated, quality machines and processors are available in the market which can handle the huge data without any problem and many organizations are using.

e. Availability of Huge Data: The data available today is huge and the time it takes to get double is unimaginable. For example YouTube, huge multimedia collection is available and each and every second it is increasing rapidly.

5. Big Data Tools & Techniques

There are various tools and techniques [6] available to handle Big Data and still researches on Big Data are taking place by information scientists and academicians across the globe. Some of the important tools and techniques used in Big Data handling are as follows:

Table 2

Tools	a. NoSQL: Databsed Mongo DB, Couch DB, Cassandra, Redis, Big Table, Hbase, Hypertable, Voldemort, Riak, and Zookeeper etc.
	b. Map Reduce: Hadoop, Hive, Pig, Cascading, Cascalog, Mrjob, Caffeine, S4, Map R. Acunu, Flume, Kafka, Azkaban, Oozie, and Greenplum etc.
Data Analysis Techniques	a. Storage: S4, Hadoop Distributed File System etc.
	b. Servers: EC2, Google App Engine, Elastic, Beanstalk, Heroku etc.
	c. Processing: R, Yahoo!Pipes, Mechanical Turk, Solr/Lucene, Elastic Search, Datemeer, Big Sheets, Tinkerpop etc.
Big Data Application	a. Social computing,
	b. Bio-Medicinal,
	c. Finance,
	d. Astronomical and so on.

6. Various Issues relating to Big Data

There may be several issues involved in big data, some of the major issues which requires the prime concern may be listed as follows:

- **a. Storage Issues:** As the data quantity is increasing day-by-day, so the storage of this huge data concerning the major issue with big data.
- **b. Transport Issues:** Now, if we have to process the data then transition of this huge data becoming a cumbersome issue.
- **c. Management Issues:** Management of big data will become the difficult problem if this rate of data remains the same.
- **d. Processing Issues:** Efficient processing of this Exabyte of data mainly need the parallel processing with advanced analytics algorithms which will give timely and effective information.

7. Risks and Challenges in big Data

If we discuss about Big Data and its related issues, be careful that there are a number of challenges that need to be addressed for you to be successful in Big Data analytics. There are various challenges involved in Big Data which need to be taken care in the process, some of the challenges [7] are as follows:

- **a. Data Integration:** The ability to combine and to do so quickly and at reasonable cost. with such variety, a related challenge is how to manage and control data quality so that you can meaningfully connect well-understood data from your data warehouse with data that is less well understood.
- **b. Data Volume:** The ability to process the volume at an acceptable speed so that the information is available to decision makers when they need it.
- **c. Skills Availability:** Big Data is being harnessed with new tools and is being looked at in different ways. There a shortage of people with the skills to being together the data, analyze it and publish the results or conclusions.
- **d. Timeliness:** It is directly proportional to size, more time will be required to process and analyze data as the size being larger. the effective system is that which provide data to the user in correct form and also on right time.
- **e. Personal Privacy:** In this modern epoch of Big Data where personal information which is stored as well as transmitted via ISPs, mobile network operators, local councils, supermarkets, financial and medical service organizations. Also through social media i.e. Twitter, Facebook etc. information stored and shared at these sites, privacy plays an important issue for everybody. Everyone wants to conceal their delicate data in order to keep away from the misuse of this particular information. But as the Data is increasing, it is very problematic to attain security of this Big Data.
- **f. Solution Cost:** Since Big Data has opened up a world of possible business improvements, there is a great deal of experimentation and discovery taking place to determine the patterns that matter and the insights that turn to value.
- **g. Performance:** Performance is an important aspect involved in any system and the popularity and survival depends upon the performance of the system. Performance become more important as the data grows faster than energy on chip, efficiency and scalability etc.
- **h. Effectiveness:** It is also a great matter of concerns that up to what extent of effectiveness system works.
- **i. Flexibility:** The system involved in Big Data is also somehow a big challenge. It is expected that the system should be flexible enough to include new and important things and to exclude the outdated and useless things.

8. Principles for Designing Big Data Systems

To develop an effective and efficient Big Data system there are some of the principles [8] based on the various researches output and experience of research scholars can be listed as follows:

- **a. Good architectures and frameworks are necessary and on the top priority:** there should be good and a well-structured architecture to efficiently and approvingly solve Big Data. However, traditional system architectures cannot be used for the analysis of the Big Data, which require high-speed systems. There is a need of parallel processing and distributed architectures for the problems related to the Big Data.
- **b. Support a variety of analytical methods:** As the data created by the Big Data is so complex to work with, and using only one or few disciplines and with counted number of analytical methods doesn't make any sense for its variant applications. In this modern science of data, there is constantly involvement of interdisciplinary approaches, which ranges from machine learning, statistical analysis, data mining and visualization and distributed programming to real-time and in-memory

analysis and computer-human interaction. These methods collaboratively applied to many Big Data platforms.

c. **No size fits all:** when we talked about Big Data analytics, it is not possible to have one general size that can fit for ball solutions, according to IBM's Latin America Big Data sales leader, Leonardo. As we aware that every tool has its own limitations, but if an individual uses the proper set of tools for other different tasks, they can also in some manner obtain a significant benefits via using these tools. In the near future, Big Data problem will surely be converted into the small data problem, as the data keeps growing at exponential rate.
d. **Bring the analysis to data:** As the time passes, the size of big data is converted to Exabyte; it becomes infeasible to accumulate and transit this data for analysis to one or several different centers. Data-driven based analysis needs different analysis direction, which requires bringing the analysis jobs to data sites, whereas for data-intensive computation problems, data is not analytical human or machine, but is driver.
e. **Processing must be distributable for in-memory computation:** in-memory analytic, traditionally, data must store in RAM not on disk is most popular as it increases the process of analysis, even if the volume of data explodes. This analytic is also importantly necessary for the real-time analytic. Thinking of this, we consider that applications which are based on real-time will be benefitted from this type of analytic process.
f. **Data storage must be distributable for in-memory storage:** many problems related to the Big Data involve mainly data and the useful information is created at different times and at variant addresses, this point is met already. In the case where generated data accumulated at a data center, partitioning of the data has to be done for the in-memory analytic approach. Cloud has been used as the space for data storage. Once the data has been stored, users start their calculation using this Big Data on powerful computers.
g. **Coordination is needed between processing and data units:** Coordination between various processing and data units is importantly essential for the fault-tolerance, efficiency and scalability of Big Data systems. The low latency of response which is mainly needed in real-time analytics is guarantees by this principle.

9. Conclusion

As the concept big data is getting its momentum across the globe, the professionals, experts and scholars have already started taking keen interest to understand its impact and applications. In short, it can be concluded that big data is a sea change that, like Robotics, Nanotechnology and Quantum Computing, will shape the twenty first century [9]. Through efficient analysis of the Exabyte data which is becoming available, there must be potential for the creation of faster advances in various disciplines and civilizing profitability and triumph of several enterprises. However, many technical and non-technical challenges discussed in this particular article must be addressed before this prospective can be realized fully [10].

10. References

1. Quora. Retrieved on from https://www.quora.com/Who-coined-the-term-big-data, 2017.
2. Najafabadi, Maryam M, and others. Deep learning applications and challenges in Big Data analytics. Journal of Big Data. 2015; 2(1):1-21.
3. Ranjana, Bahri, Big Data Concept, Challenges and Management Tools. International Journal of Advanced Research in Computer Science and Software Engineering. 2015; 5(2):986-990.
4. Oguntimilehin A, Ademola EO. A Review of Big Data Management, Benefits and Challenges. Department of Computer Science, AfeBabalola University, Ado-Ekiti, Nigeria.
5. Riccardo Torlone, Big Data: An Introduction, Retrieved on from http://www.dia.uniroma3.it/~torlone/bigdata/L1-Introduzione.pdf, 2017.
6. Marko, Grobelnik. Big Data Tutorial. Retrieved from http://www.planet-data.eu/sites/default/files/presentations/Big_Data_Tutorial_part4.pdf, 2017.
7. Lavastorm Analytics. The Top Challenges in Big Data Analytics. Retrieved from http://www.gsma.com/membership/wp-content/uploads/2013/07/The-Top-Challenges-of-Big-Data-Analytics.pdf, 2017.
8. Chen CL. Philip and Zhang, Chun-Yang. "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Science. 2014; 27(5):314-347.
9. Global Pulse. Big Data for Development: Challenges and Opportunities. Retrieved. From <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf, 2017.
10. United States. Challenges and Opportunities with Big Data. Retrieved from <http://cra.org/ccc/wp-content/uploads/sites/2/2015/05/bigdatawhitepaper.pdf, 2017.