A study of scope and impact of Big Data analytics on Information Technology in India

Mr. Shaikh Mohammad Ali[1], Mr. Shoaib Sayyed[2].

MBA IT [1], Operations [2] Poona Institute of Management Sciences & Entrepeneurship[1,2]
Savitribai Phule Pune University, Pune, Maharashtra, India[1,2]
Email: mohammadali.pimse12@gmail.com[1] , shoaibsayyed20@gmail.com [2]

Abstract- Big data is leading topic across a various fields, with both the sectors having use of big data analytics. It is determine by size, type, flow of data, and by the using various algorithms data. In this paper, understand the term 'big data'. The role of big data analytics does not involve personal data. Using climate and weather data. Big data analytics can involve repurposing personal data. Any organization has collected some personal data for commercial purpose and then decides to start analyzing it for completely different purposes then it needs to make its users aware of this. This is particularly important if the organization is planning to use the data for a purpose that is not apparent to the individual because it is not obviously connected with their use of a service.

A key feature of big data is using 'all' the data, which contrasts with the concept of data minimization in the data protection principles. This raises questions about whether big data is excessive, while the variety of data sources often used in the analysis may also prompt questions over whether the personal information being used is relevant. The challenge for organizations is to address this by being clear from the outset what they expect to learn or be able to do by processing that data, as well as satisfying themselves that the data is as per requirement or not. Organizations must have security issues of any information stored at big data. Security depends upon the proper assessment of risk, and so responsible organizations should apply their normal risk management policies and procedures when they acquire new datasets or use existing ones for big data analytics. Big data can also be a tool to improve information security. As well as looking extensively at the data protection issues presented by big data, this paper also suggests areas   .

Index Terms-   Data Protection Act [DPA], Small and Medium enterprises [SME's], Online Analytical Programming [OLAP], Centre for Economics and Business Research [CEBR], Complex Event Processors [CEP], Data Stream Management Systems [DSMS], Big Data Analytics Agents [BDAA].

1.   INTRODUCTION:

1.1 Definition **: "**Big Data" is a term a direction which uses for various techniques to take input data, process, and analysis and access large amount Data for specific time set frame not accessible to various standard IT technologies. But using bridges, and necessary platform, tools, software etc. used for necessary requirement are together form called "Big Data technologies". The new perspective of Big Data is dealing with the maintenance cost of storing and processing of Data the cost of storage has been reduced by many segment by the Cloud computing, which cuts huge amount of economy of organization by dealing with various   IT investment cost for all business. As a result the "Big Data concept" has   opportunity   for   many   business   & developing Research centers for various objective.
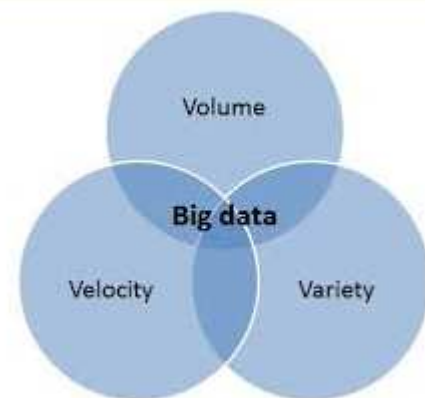

Fig 1. Big data concept

1.2   Origin of the concept:
A few decades before, data storage was one of the major issues in for business. Now new standards, efficient, scalable and tangible upcoming technologies have integrated with Database   System   for   secured   storage resolving the data storage issues. Resulting in frequent access of data either by internet or Offline, data is constantly being generated, and also large amount of information is gathered from various sources like automated system etc. This phenomenon has recently accelerated due to increase of connected devices and the worldwide success of the

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

social platforms. Major social platform like Google, Wikipedia and Face book and Twitter, has enhanced the application of data

an individual, it is necessary to consider what means are reasonably likely to be used to identify them. The ICO has produced guidance that explains this
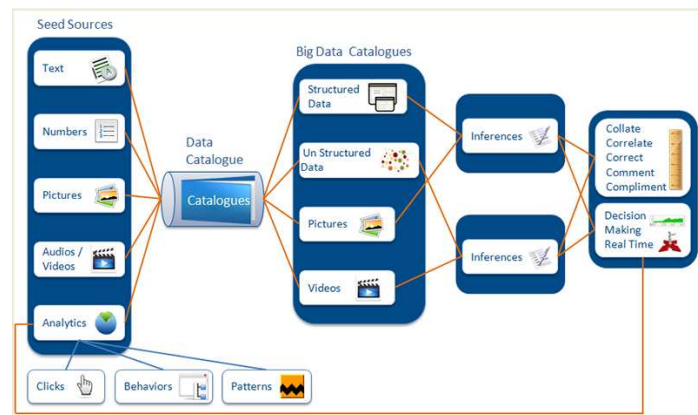
> ➢ **Fig no. 2. Work flow of Big Data.**

volumes at internet on large scale depend upon situation .Now these solutions have partly moved into the some open source and easily available to all. It led the approach to Big Data with easy and cost free solution for business problems.

➢ Meanwhile, two parallel breakthroughs have further helped accelerate the adoption of solutions for handling Big Data:

➢ The availability of Cloud based solutions has dramatically lowered the cost of storage, amplified by the use of commodity hardware. Virtual file systems, either open source or vendor specific, helped transition from a managed infrastructure to a service based approach;

➢ In term of large volume of data over the server is distributed data and load balancing over many servers. Up gradation of database design for storage and Parallel or Multi processing has given support to various Business products originating into Hadoop platform. Upon acceptance, authors are required to submit their data source file including postscript files for figures.

1.3 Big data and personal data:
Personal data is the data relates to a living individual who can be detected. 'Identify' means individual can be identified from the data, either individual or combined information. In assessing whether the data could be combined with other information to identify

definition further Data protection is concerned with personal data, but it is important to remember that many instances of big data analytics do not involve personal data at all. E.g. of non personal big data include: world climate and weather data; using geospatial data from GPS-equipped buses to predict arrival times; data from radio telescopes in the Square Kilometer Array. Data from sensors on containers carried on ships. These are all areas where big data analytics enable new discoveries and improve services and business processes, without using personal data. Big data analytics that do involve processing personal data, for example: Monitoring devices of data in Medical, Mobile, and Location with respect to data, with respect to purchase data. Big data analytics has potential to generate new personal data. For example, social media and other data about an individual could be analyzed to find the person's lifestyle as a factor in determining credit rate, or risk during of developing a medical condition. Car pooling system has various sensors in cars provide all data about the car as well as individuals for various decisions making, Big data can be used to understand the emerging and current trends for decision making. For example, data from travel cards that record journeys made by individuals, such as Oyster card, could be combined with traffic data to plan new bus routes. In such cases the data may often be anonymised before being analyzed.

2. SCOPE AND IMPACT OF BIG DATA A NEW WORLD OF OPPORTUNITIES:
According to the McKinsey Global Institute report on Big Data from 2012, the most developed regions, such as Europe, have the biggest potential to create value through the use of Big Data. The enormous economic impact of Big Data is further shown in another study prepared by the Centre for Economics and Business Research (CEBR),

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

estimating the value of Big Data to the UK economy alone, being £216 billion and 58,000 jobs in the next 5 years. Big Data is further expected to add more than €250 billion a year to the European public sector administration. Thus, the whole European Union could benefit from the cumulative financial and social impact of Big Data. Big Data analytics has started to good impact on all organizations, as it can dig out huge amount of data to for Knowledge and real time application. These may results in development of "smart cities", where data is collected by various communicable systems and find out the transactions. The limitation lies in security of data as applications is on open source depending upon demand and betterment of citizens in various approaches of E-Business .

Various branches of medical science generate large amount of data. Big Data in Medical transcription is associated with data pool of patient. Example In medical care in magnetic resonance imaging and in CT scans.. Doctors usually require the support of automated solutions, which is based on the application of machine learning.

> Big Data technology is required for various case studies in mobile networks. Big Data is important for example for managing and operating mobile networks and goal to improve the network quality; which includes isolation and faults within the network, support of security related detection and prevention mechanisms, traffic planning, prediction of hardware maintenance, calculation of drop call probability.

> The changes brought by new web social media in the area of news media, conventional journalism etc., operating with standard news collection and broadcasting procedures while mediating mainstream types of content from authoritative sources. Traditional news media are replaced by web news services.

> In terms of business and economic activity, many software and services providers rely on Online Analytical Programming (OLAP) systems to perform their market analysis. Big Data technologies help scale in legacy systems. For other approach business is also supported by Big Data technologies.

3. Big Data Analytics :

There is a huge development in areas like machine language, Encryption etc. Therefore, the scope is also

overcome the requirements of various standards in data mining techniques:

- a ecosystem may be created to adopt the various method or design
- a new steps in algorithm
- a technology platform and adequate development skills to be able to implement it
- Data structure of the system, for various business value.

Giving rise to various researches Methodology of data for various development. As data is said to be valuable asset to organization for business perspective, but due lack of understanding of analyst for improve the business policy. As per the object design the modeling of data is complex and large in distribution. As per various software solutions, the niche developed is general, compatible with respect to data visualization.

 Some of the factors required in Big Data analytics are as follows:

i. Context awareness

ii. Rethinking data visualization and human-computer interfaces

iii. Visual Analytics: how we look at data

iv. Data management performance and scalability

v. Correlation and Causality

vi. Real time analytics and stream processing

vii. Distributed Storage

viii. Content Validation

 Visualization is the means interaction between users and computers for various applications.

The features of visual analytics research are as follows:

i. Emphasing on data for analysis, problem solving, along with decision making;

ii. Expert system for data processing and new steps for approaching techniques for meeting the requirement

iii. User friendly Interface for the end users

iv. Report generation for any time period

v. Decision support system for analysis of proper data.

Performance and scalability are has got some technical issues with large volume of Data to be stored and processed by Big Data systems and technologies. Currently, process analysis in areas such as Business process management (BPM) is adapted with type of data to be analyzing in areas such as Data Mining and Business Intelligence. Bridges focusing on process mining will be also essential for progression of the Big Data theme. Therefore, any Big Data strategy needs to focus on a few areas:

• Discovering and modeling causality in structured data

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

• Discovering and modeling causality in unstructured
• Integrating unstructured causality models with structured causality models
• Approaches based on traditional solutions like Data Stream Management Systems (DSMS) and Complex Event Processors (CEP), are generally insufficient for the challenges posed by stream processing in a Big Data context: the analytical tasks required by some analysis to generate the knowledge automatically reasoning tasks are also needed. New Big Data-specific parallelization techniques and (at least partially) automated distribution of tasks over clusters are crucial elements for effective stream processing. Achieving industrial grade products will require:
• New techniques to associate quality preferences/ Requirements to different tasks and to their interworking relationships;
• New frameworks and open APIs for the quality-aware distribution of stream processing tasks, with minimal development effort requested by application developers and domain experts.

The approach is the use of Big Data Analytics Agents (BDAA). The agents perform a specific function or set of functions on data sets. They are dispatched to appropriate locations and must be security cleared by the receiving data set. Once they have performed their analytics the data set then verifies the agents' findings before letting it return to its sending location. BDAA's could then be designed for the specific features of Big Data .Validating the vast amount of information in content networks is a major challenge, since there is a very large number of different types of sources, such as blogs, social networking platforms, or news sites with social networking functionalities, and different types of content, such as articles, comments, tweets, etc. Furthermore, the complexity of human language is such that it is not easy to derive validation rules that apply to all different discussion subjects.

Davenport ETal25 state that organizations capitalizing on Big Data differ from traditional data analysis in three ways:
• They pay attention to data flows as opposed to stocks.
• They rely on data scientists and product and process developers rather than data analysts.
• Analyst are now switching themselves from IT to various branches such as HR, Finance , Production etc as key areas   **ref[2].**

4. Table for Strength and weakness for SME :

Sr. no	Strength	Weakness
1	Globe has a large group of commercially successful SME's focusing on market driven innovation within their niche and with a focus on export. In Germany, they are known as "hidden champions" but they exist in other European countries as well. Because these SME's operate in very narrow niches they have to focus on global markets to work on an economic scale. Those companies benefit from exclusive insights about their global customers to build market oriented products and services.	While US-based companies like Yahoo, Google, or Twitter are widely recognized for their activities in Big Data, very few research organizations, including SMEs, are known for their activities and initiatives in this field in Globally
2	There is deep knowledge about local markets and local customer problems and ability to develop customized products such as language dependent products, legislation dependent products.	SME's lag behind larger enterprises in taking up the Big Data challenges, especially compared to the US In most SME's there is not enough understanding on how to gain new insights by using data analytics concerning their customers, products and services.
3	Effective (although fragmented) research and development networks between universities, research centre's and SME's are established and has been referred to as a "role model for decentralized R&D".	The capabilities to develop an individual data strategy, to select and integrate the right data sources and to use effectively big data analytics for leveraging exclusive insights for product and business development are

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

		critically underdeveloped.

iii. Globally SME's could benefit enormously by reducing copyright infringements with smart networked products and services.
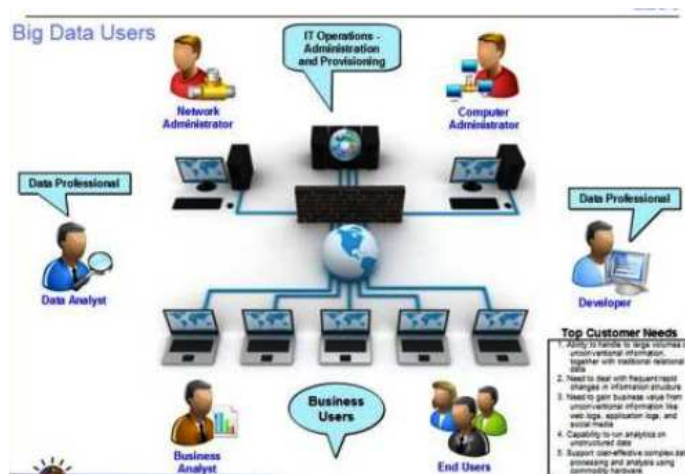
> ➤ **Fig no. 3. Big Data user.**

4	There is a well-established understanding in export oriented companies to deliver high-value products and services (e.g. "premium cars") by using innovation networks	No wide spread knowledge about freely available data ("getting new insights by merging different data sources")
5	Growing interest from SME's using cloud computing and software services	Analytical Big Data services for SME's within Europe are currently non-existing

.

5. Opportunities of Big Data:
 i. A rising customer demands in Europe for smarter products, higher individualization, and mass customization.
  ii. Globally SME's can benefit by enhancing their products and services with big data analytics and privacy by design, e.g. offering preventive maintenance services in the utilities industry, usage-based analytics for distant product development, self-learning behaviors for energy optimization (e.g. Nest thermo state) or new business models thanks to better usage insights (like Rolls-Royce and their power by the hour performance based contracting).

  iv. Better use of freely available data.
  v. Leveling the playing field by giving access to formerly very demanding analytical tools through commoditization.
  vi. Developing new products and services enhanced with Big Data analytics and privacy by design, developing products adapted to European privacy standards
  vii. Established innovation networks could reduce skill shortages by spreading knowledge with online trainings and generating hands-on expertise based on commoditized analytical services.

6. Threats of Big data:
  i. Globally there is a fast growing knowledge about using data as an asset class for leveraging the industrial competitiveness, for example in China, India.
  ii. There is a massive venture-capital driven development in the US to commoditize formerly complex big data analytics for the mainstream market, www.bigml.com is an example of the commoditization efforts around Hadoop. SME's in Europe have no alternatives to choose comparable services from European vendors.
  iii. Insufficient capabilities of European companies to scale to the world market.
  iv. There are serious concerns among European businesses about using big data analytics

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

neglecting European privacy standards and expectations. In the meantime, foreign competitors take their

available in open source, for faster growth of Business to increase profitability However, many technical challenges described in this paper must be addressed
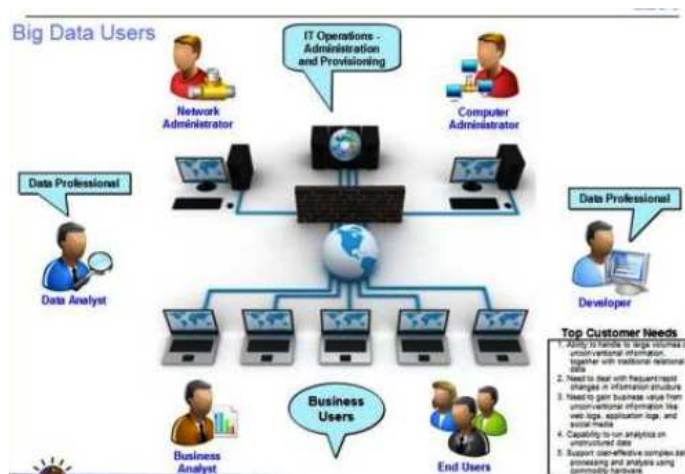
> ➤ **Fig no. 3. Big Data user.**

7. Technical aspects:

It purely deals with Hardware & Software specification and with respect to real time Implementation is the big challenges in the real time Environment, data management and performance and scalability, pros and cons to some extent, distributed storage **ref.[2].** In detail, the following directions for future research efforts.

1. The methodology towards data analysis with quick approaches algorithms by generated massive data collections (heterogeneity, multimodality, size etc.)

2. The probability of data to used for betterment of the Business.

3. With regards to visual analytics, it is important to - on a general scale - promote the need for understanding the relevance and relatedness of information.

4. There is requirement for the presentation on the type of data it may temporary or permanent scales. Therefore necessary tools supporting various platform interactions with user friendly transitions for aggregation of data to one another. ref.[2].

8. Conclusion:

We have entered an era of Big Data for better analysis of the large amount of data size which are

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. The technical challenges are common across the globe, and are not cost-effective to address in the context of one domain alone. Furthermore, these challenges will require alternatives or parallel solutions, and will not have impact on next generation of business products. We must support and encourage fundamental research towards addressing these technical challenges if we are to achieve the promised benefits of Big Data.

9. REFERENCES:

[1] Big data and data protection 20140728 Version: 1.0
[2] NESSI White Paper, December 2012 Big Data A New World of Opportunities
[3] https://ico.org.uk
[4] http://www.ijircst.org
[5] https://www.linkedin.com
[6] http://its.fhsalzburg.ac.at - Koehler Martin
[7] http://sloanreview.mit.edu

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

Results companies have achieved using big data analytics

➤ **Fig no. 4. Big Data Profitability**