International Journal of Research
eISSN: 2348-6848 & pISSN: 2348-795X Vol-5 Special Issue-13
International Conference on Innovation and Research in Engineering, Science & Technology
Held on 23rd & 24th February 2018, Organized by Tulsiramji Gaikwad Patil College of Engineering & Technology, Nagpur, 441108, Maharastra, India.

# Analysis of Big Data

**Miss: Shivani Dnyaneshwar Nagpure, Miss: Pallavi Chedge,**
M.C.A. VI Semester, G.H. Raisoni College of Engineering Nagpur, Maharashtra.

*Abstract— Big data is a popular, but poorly defined marketing buzzword, that describe the exponential growth, availability and use of information, both structured and unstructured. The big data trends and how it can serve as the basis for innovation, differentiation and growth. Looking at big data is that it represents the large and rapidly growing volume of information that is mostly untapped by existing analytical application and data warehousing systems. Example data include high-volume sensors data and social networking information form web sites such as Google, Face book, LinkedIn, Yahoo, Amazon and Twitter. The exponential growth in the amount of biological data needed for data management, analysis and responsibility..*

***Keywords***

*Social Media, Structured data, unstructured data, Hadoop, NoSQL, Data Warehousing Things.*

## 1. Introduction

Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tool or traditional data processing applications. Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the structure of yours database architectures. To gain value from this data, you must choose an alternative way to process it. The Value of big data to an organization falls into two categories: analytical use, and enabling new products. Big data analytics can reveal insights hidden previously by data too costly to process, such as peer influence among customers, revealed by analyzing shoppers' transactions, social and geographical data. Being able to process every item of data in reasonable time removes the troublesome need for sampling and promotes an investigate approach to data, in contrast to the somewhat static nature of running predetermined report

### 1.1 Types of Data

Big Data is made of structured and unstructured information. The term structured data and unstructured data refers to that is identifiable based on it is organized in a structure or not Collection and analysis of data far beyond the scale of what was previously possible.

Fourth, the mobile computing cloud enables developers to offload mobile services to back-end servers, providing unprecedented scale and additional resources for computing on collections of large-scale sensor data and supporting advanced features such as persuasive user feedback based on the analysis of big sensor data.

In this article we give an overview of the sensors on the phone and their potential uses. We discuss a number of leading application areas and sensing paradigms that have emerged in the literature.

#### 1.1.1 Structured data

The most common form of structured data or structured record is a database where specific information is stored based on a methodology of columns and rows. Structured data is also searchable by data type within content. Structured data is understood by computers and is also efficiently organized for human readers. Relational database and spreadsheet are example of structured data. In contrast, unstructured data has no identifiable structure. Structured information is the data in databases and is about 10% of the story

#### 1.1.2 Unstructured data

The term unstructured data refers to information that either does not have a pre-defined data model and/or does not fit well into relational tables. The terms unstructured data refers to any data that has no identifiable structure. Unstructured information is 90% of Big data and is 'human information' like emails, videos. Tweets, Face book posts, call centre conversation, closed circuit TV footage, mobile phone calls, website clicks. Big Data is only getting bigger 90% of the data in the world today was created within the last two years. While each individual document may contain its own specific structure or formatting that based on the software program used to create the data unstructured data may also be considered "loosely structured data" because the data source do have a structure but all data within a dataset will not contain the same

structure but all data within a dataset will not contain the same structure. Sensors. These are probably the most ubiquitous sensors on the planet. By continuously collecting audio from the phone's microphone, for example, it is possible to classify a diverse set of distinctive sounds associated with a particular context or activity in a person's life, such as using an automatic teller machine (ATM), being in a particular coffee shop, having a conversation, listening to music, making coffee, and driving. The camera on the phone can be used for many things including traditional tasks such as photo blogging to more specialized sensing activities such as tracking the user's eye movement a cross the phone's display as a means to activate applications using the camera mounted on the front of the phone. The combination of accelerometer data and a stream of location estimates from the GPS can recognize the mode of transportation of a user, such as using a bike or car or taking a bus or the subway. More and more sensors are being incorporate into phones. An interesting question is what new sensors are we likely to see over the next few years? Non-phone-based mobile sensing devices such as the Intel/University of Washington Mobile Sensing Platform (MSP) have shown value from using other sensors not found in phones today (e.g., barometer, temperature, humidity sensors) for activity recognition; for example, the accelerometer and barometer make it easy to identify not only when someone is walking, but when they are climbing stairs and in which direction. Other researchers have studied air quality and pollution using specialized project or the Mobile Millennium project (a joint initiative between Nokia, NAVTEQ, and the University of California at Berkeley) are being used to provide fine-grained traffic information on a large scale using mobile phones that facilitate services such as accurate travel time estimation for improving commute planning.[2][8].

## 2. WHAT DOES BIG DATA LOOK LIKE?

In 2012, Gartner formalized their Big Data definition as a "3V" framework – high Volume, high Velocity, and high Variety information asset, requiring new forms of processing to enable enhanced decision making, insight discovery and process optimization. The IBM adds a fourth "V" of Veracity to add trust and noise filtering to the challenge of Big Data analysis. Input data to big data systems could be chatter from social networks, web servers logs, traffic flow sensors, satellite imagery, broadcast audio streams, banking transaction, MP3s of rock music, the content of web pages, scans of government documents ,GPS trails, telemetry from automobiles, financial market data, the list goes on .

### 2.1 Volume

There are many factors contribute to the increase in data volume- transaction-based data stored through years, text data constantly streaming in from social media, increasing amounts of sensor data being collected, etc. In the past, excessive data volume created a storage issues, But with today's decreasing storage costs, other issues emerge, including how to determine relevance among the large volumes of data and how to create value from data that is relevant. This volume presents the most immediate challenge to conventional approach to querying. Many companies already have large amount of archived data, perhaps in the form of logs, but not the capacity to process it.

### 2.2 Variety

Big Data is any type of data; today data comes in all types of formats- from traditional database to hierarchical data stores created by end users and OLAP systems, to text documents, email, meter-collected data, video, audio, stock ticker data and financial transactions. By some estimates, 80 percents of an organization's data is not numeric. But it still must include in analysis and decision making. Different browser send different data users withhold information, and they may be using differing software versions or vendors to communicate with you. A common use of big data processing is to take unstructured data extract ordered meaning, for consumption either by humans or as a structured input to an application.

### 2.3 Velocity

Velocity "means both how to fast data is being produced and how fast the data must be processed to meet demand". Problem previously restricted to segments of industry are now presenting them in a much broader setting. Specialized companies such as financial traders have long turned systems that strive with fast moving data to their advantage. The Internet and mobile era generating a data flow back to provider. The Retailers who are able to quickly utilize that information, by recommending additional purchase, for instance, gain competitive advantage. The Smartphone era increase again the rate of data inflow, as consumers carry with them a streaming source of reallocated images and audio data.

converging view and move the field forward.[3][8]

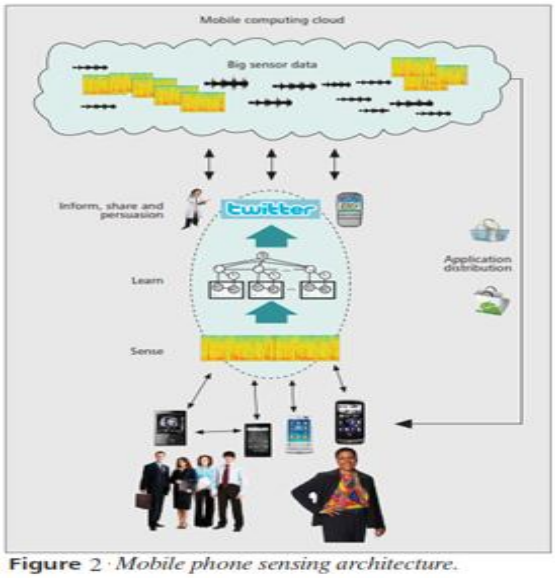


**Figure 2** *Mobile phone sensing architecture.*

### 2.4 Veracity

Veracity deals with uncertain or impressive data. In traditional data warehouse there was always the assumption that the data is curtained, clean and precise. That is why so much time was spent on Master Data Management, Metadata management, Identity Insight/ Assertion, etc. However, when we start talking about social media data likes Tweets, face book posts, etc. how much faith should we put in the data. This data can be used as a count toward your sentiments, but you would not count it toward your total sales and report on that. Due to the velocity of data likes stock trades, machine/sensors generated events, you cannot spends the time.
LEARN Information is extracted from the sensor data by applying machine learning and data mining techniques. These operations occur either directly on the phone, in the mobile cloud, or with some partitioning between the phone and cloud. Where these components run could be governed by various architectural considerations, such as privacy, providing user real-time feedback, reducing communication cost between the phone and cloud, available computing resources, and sensor fusion requirements. We therefore consider where these components run to be an open issue that requires research. [8]

## 3. BIG DATA AND BIGGER OPPORTUNITY

Even though "Big Data" has now been around for a few years, the opportunities for start-ups seem to keep growing, just as the amount of data keeps growing. According to IBM, companies have captured more data in the last two years than in previous 2000 years. This data comes from sensors, social media posts, digital picture and video, purchase transactions, everywhere. Every day, we create 2.5 quintillion bytes of data, much of it unstructured and far beyond the capability of conventional databases. Hence one segments of the opportunity is the need for new database technologies, like Hadoop, a distributed file system originally designed for indexing the web. Data capacity is measured in petabytes (100 terabytes), or soon even yottobytes ($10^{24}$).

## 4. Conclusion

New systems using big data will extend, and possibly replace, our traditional DBMS's. There is no questions that there is enough data available that traditional database management systems will be defeat completely. Moving forward with big data systems, and that the best way is to start small and prove the benefits. While this is not much different from any other new technology, it might be an especially good strategy to apply to big data application. Cloud computing may also prove valuable for big data. Currently available systems for health care domains and social media and retail limits the functionalities due to the ever increases in demands. As a result the integration of new technologies is necessary to cope up with on-demand. In the future, further studies will be conducted for improving the big data platforms through theory and experiments.

of Smart Phone sensors have the potential to revolutionize various fields of human life. In-built mobile phone sensors have many such capabilities that can improve people's lives cutting down the time it takes to find things, to prevent people from getting lost, improve health conditions, and even more serious applications are emerging that could actually save lives. Security and privacy is one of the utmost issue that needs more attention while developing mobile phone sensing systems as when Mobile phone is used for social interactions users main concern is to secure their private data. However, this study highlights the fact that there are still a lot of challenges and issues that need to be resolved for mobile phone sensing systems to become more applicable to real-life situations and spawn further research in this area.[3]

REFERENCES

[1] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, Marimuthu Palaniswami: "Internet

of Things (IoT): A vision,architectural elements, and future directions."Future Generation Computer Systems 29 (2013) 1645-1660.

[2] G. Anilgoud, K. Morarjee: "Mobile Phone Sensing- A Review." Vol. 2, Iss. 10, Oct. 2012. International Journal of advanced research in computer science and software engineering.

[3] Wazir Zada Khan, Yang Xiang.: "Mobile Phone Sensing Systems: A survey." IEEE Communication Surveys and Tutorials. Vol. 15, No. 1, First Quarter 2013.

[4] Esther Ezra, Jiemin Zeng, Jie Gao.: " A Distributed Algorithm for Approximate Mobile Sensor Coverage." CCCG 2014. Halifax, Nova Scotia, Aug. 11 13 2014.

[5] Thejaswini M, P. Rajalakshmi, and Uday B. Desai: " Novel Sampling Algorithm for Human Mobility-Based Mobile Phone sensing." IEEE Internet of things Journal, Vol. 2, No. 3, June 2015.

[6] Salil S. Kanhere: " Participatory Sensing: Crowdsourcing Data from Mobile Smartphones in Urban Spaces."

[7] Yun Wang, Yanping Zhang, Jiangbo Liu and Rahul Bhandari: " Coverage, Connectivity and Deployment in wireless sensor networks." Chapter 2.

[8] Nicholas D. Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, Andrew T. Campbell, Dartmouth College: " A survey of mobile phone Sensing."

[9] Yao-Chung Hsu, Chi-Han Lin, Wen-Tesuen Chen: " Design of a sensing service architecture for Internet of Things with sensor selection." 2014 IEEE 11th Intl. Conf. on Ubiquitous Intelligence & Computing.

[10] Adarsh Kumar. Krishna Gopal, Alok Aggarwal.: " Simulation and Analysis of Authentication Protoclos for Mobile Internet of Things." 2014 Intl. Conf. on Parallel, Distributed and Grid Computing.