Trends and Techniques of Handling Torrents of Big Health Data towards Enhancing Healthy Life Span

M V Bramhananda Reddy

Professor and Principal Dept of Computer, Science and Engineering Sai Rajeswari Institute of Technology, Lingapuram Proddatur

Abstract

The current trend of society generates torrents of data across various sources like social networking, health sectors, mobile sensors, industries. This voluminous data raised a scope for uncovering hidden insights of this data. This huge data often called big data could undergo several data analytics to retrieve the unnoticed patterns, trends, associations, querying, and information security. Here, in this paper we focus on health care industry towards applying analytics on the health data like EHR's, medical images, reports, sensors and transform this data to make out a meaningful outcome that helps towards diagnosis and prognosis at an early intervention which reduces the morbidity, sensitizing the adverse effects of infectious diseases[2]. We also discuss the existing mechanisms of handling health care data and its underlying effects that are to be tackled.

Introduction:

In the day to day society Internet plays a prominent role in the schedule of every individual for one or the other activity like data gathering, browsing, knowledge extraction, learning, communication, coming to communication activity the social networking sites like Face book, Twitter, LinkedIn alone generate massive amounts of data for a given day. This data raised starting from gigabytes(GB), to Terabytes(TB), Petabytes(PB), Zetabytes(ZB), Yottabytes(YB) and ranging towards Exabytes.

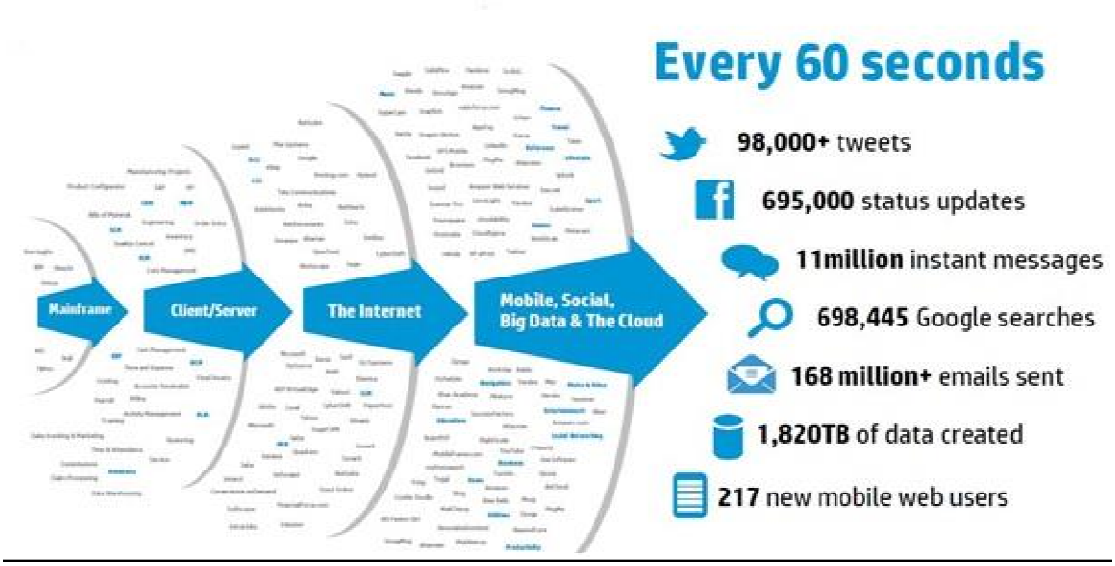

FIG 1: RELATED TO STORAGE CAPACITY

This data is not just one particular form but contains a mixture of contents like text, documents, images, audio, video etc. Storing such huge amounts of data is a major issue. Later comes the concept of information retrieval i.e. gaining meaningful insights from such voluminous data often termed as big data. Any data cannot be called Big data to be called so, it need to satisfy the basic three V's stated according to the META GROUP in 2001.The basic V's of Big data are Volume, Velocity, Variety. Volume refers to the huge amounts of data collected across different sources, Velocity refers to the speed at which the data is generated, and variety refers to the various categories of data present in the collected one. Apart from this another two V's like veracity and value are also added where veracity refers to the quality of data i.e. maintained and value refers to the originality of data.

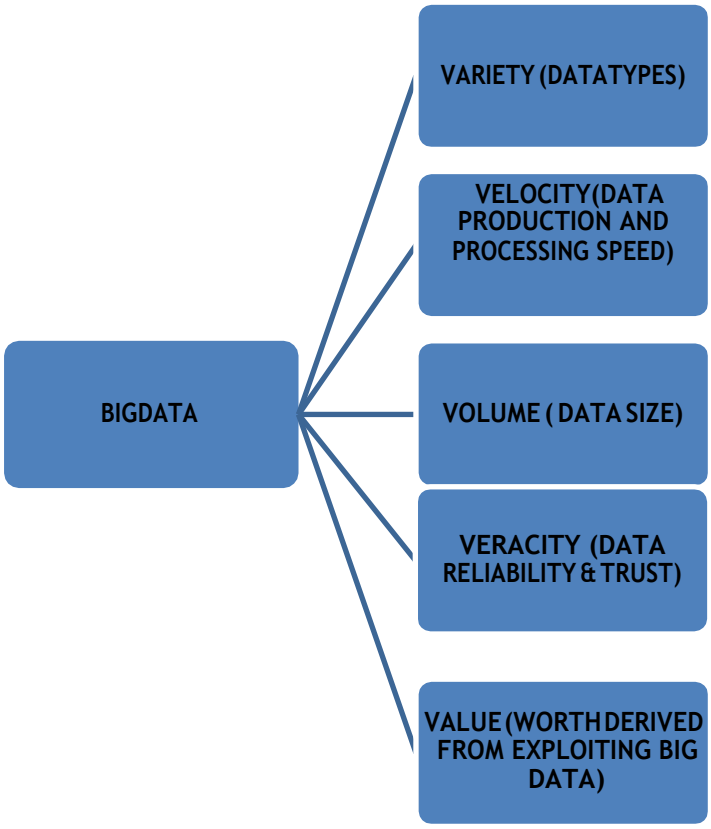

FIGURE 2: REPRESENTING THE BASIC V'S OF BIG DATA

Now a day's people spend much on two aspects of life they are luxury and health. Luxury may denote shopping, entertainment, assets, share market, business sector etc. To make a meaningful outcome of this data we generally prefer a type of analytics called Behavioural analytics, Business analytics, sentimental analytics, market analytics, risk analytics and so on. Behavioural Analytics refers to scenarios where we often see advertisements of our interested sectors during our access of web pages. This is made possible through behavioural analytics where certain features like age, gender can be considered as factors of filtering. Business analytics may refer to drawing associations of several relevant products like bread, butter, milk and eggs stored across could enhance business level, profits in a super market. Sentimental analytics may be useful to study the emotions of people collecting the

smiley's, emotions shared across chats and predict the mental status of a person based on some survey.

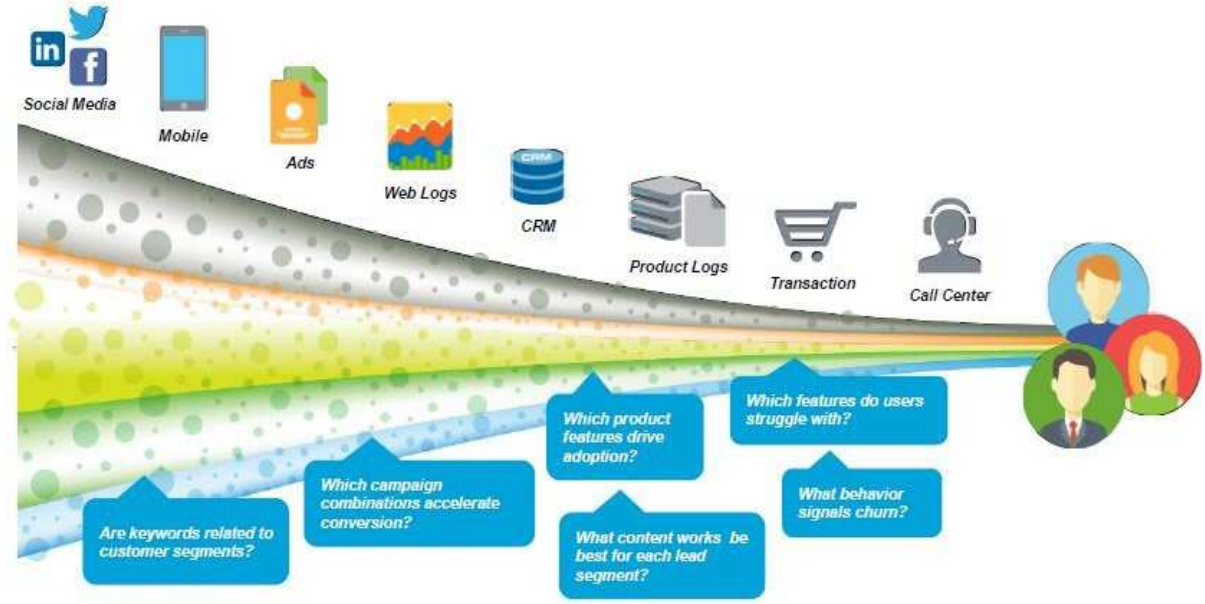

FIGURE 3 : BUSINESS OF BIG DATA

Market Analysis may refer to value of shares, trade, business etc...Risk analysis refers to predicting the level of risk indulged in the specific activity handled. Another area that requires to be focused these days is the healthcare sector. Many organisations offer several health policies, insurance to their employees and even a common man concentrates to attain several insurance schemes, towards safeguarding one self. To offer any services to its end user the health care providers gather the information of individuals from hospitals, surveys, WHO(world Health Organisation) and census data.

The health care data collected is often stored as an EHR(Electronic Health Record), which contains the minimum details of an individual like his height, weight, name, age, gender, BMP, Blood Pressure, Sugar levels etc..This health data collected across different sources may be EHR's, medical images, reports, sensor data and so on[18].

The complex environment of BIG Data consists of multiple data sources along with sophisticated analytics and multiple output forms. It consists of Data Integration, Data Management, Data Analytics and Decision Management. Deployment options are multiple for Big data and user can access through multiple devices.

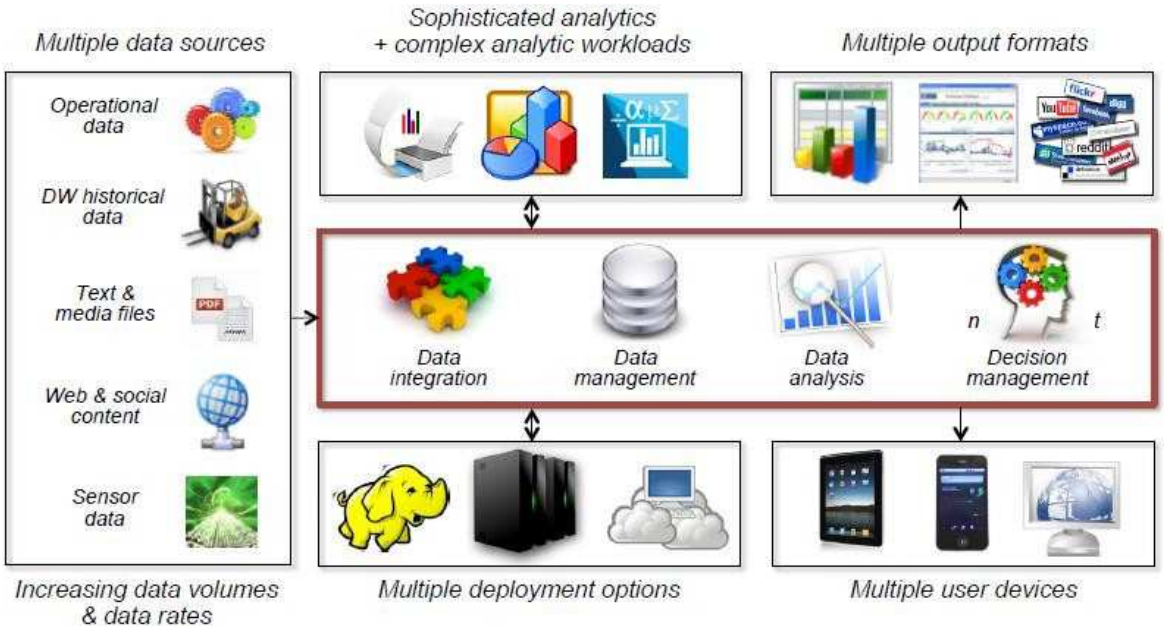

FIGURE 4: COMPLEX ENVIRONMENT OF BIG DATA

Investigative Computing for Health care Industry has Hadoop as an Example.

"If Hadoop didn't exist we would still have to make decisions about what can come into our data warehouse or the electronic medical record (and what cannot). Now we can bring everything into Hadoop, regardless of data format or speed of ingests.

This paper focuses towards Architecture of Big data, challenges, applications of big data in health industry, transforming this healthcare data to make necessary outcome, Predictive analysis, Diagnosis & prognosis, Tools and techniques used.

I. Architecture of Big data:

Handling of Big data is a challenging task since, data is generated in massive amounts at a very high velocity which is not sustainable by the traditional databases thus we look out for a new approach of handling this data, as we have not only structured data as in traditional databases but most of the real time data is semi structured or unstructured. The first task starts up with collection of data across different sources then comes the Data Storage thereafter retrieval of data to mine the necessary information to draw valuable insights. There are several platforms where Big data can be handled. In general we use a platform called Hadoop, since it's an open source, end user can program the necessary code. Hadoop contains a HDFS (Hadoop Distributed File System) which concentrates on distributed storage and fault tolerance of the data. On top of this, we use Map Reduce technique which maps similar patterns, clusters sets of data and store them[16] [17]. In the data processing we apply many machine learning algorithms to analyse the data sets collected across and stored in the data base. Not only proper storage of the data but retrieval of required information within th

specified time is a challenging task where we concentrate on Map and reduce algorithms for data accuracy and to gain value of the data

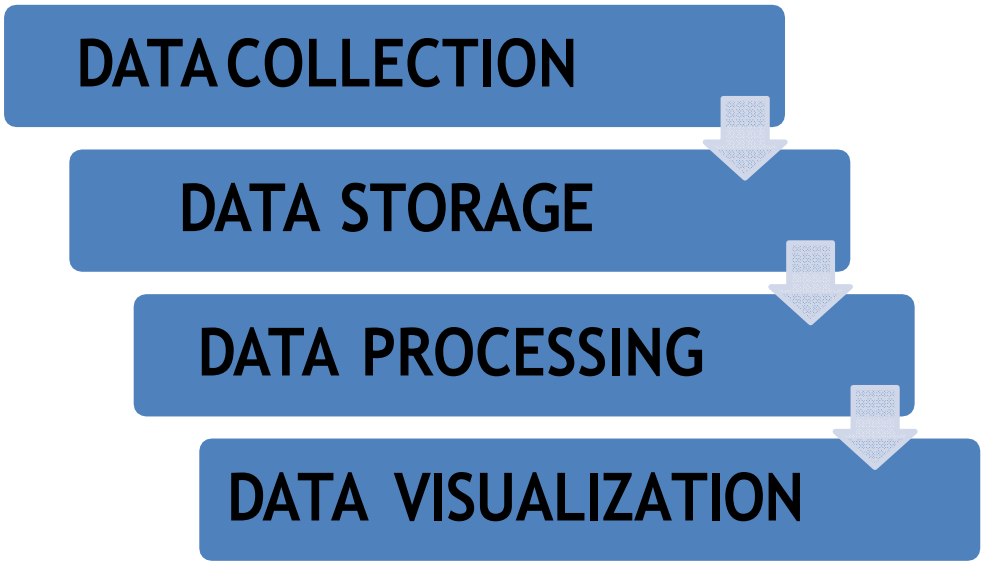

FIGURE 5: ARCHITECTURE OF BIG DATA

Health data volume is expected for a drastic growth in the coming years. Health insurance policies, reimbursement models, various life insurance policies are emerging trends of today's environment. Apart from this the data coming from various providers, hospitals stored as Electronic Health Records (EHR), Electronic Medical Records (EMR), Personal health Records (PHR) provides scope for digitizing the data, store and retrieve it to meet the purpose of single to multi physicians, individuals, providers etc..When big data is synthesized and analysed the aforementioned patterns, associations and trends can be revealed. The application of bigdata analytics in healthcare has potential benefits like early intervention, evidence based medicine, prevention and optimal management that helps in enhancing individual and population health by precautionary measures drawn using predictive analytics

II. Transforming Big health data:

There are various steps involved in transforming big data to analytics:

1. Data Collection
2. Data Storage
3. Data Processing
4. Data Visualization

   The data collection process involves collecting healthcare units of data across various available sources. This data has high heterogeneity since each hospital or organization follows up a different set of approach in projecting their data. In general the healthcare data gathered can be used for health surveillance, predictive modelling, early interventions, optimal disease management & appropriate treatment (Prognostics). Google Flu Trends has collected keywords often used in the search

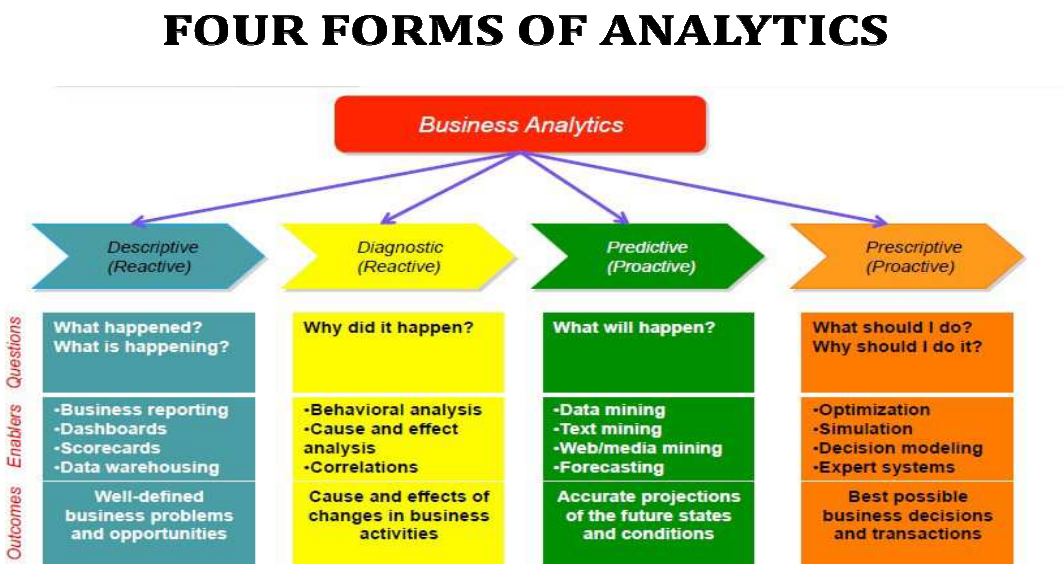

FIGURE 6: ANALYTICS OF BIG DATA

Data analysis always plays a prominent role in the market; the stored data is drilled to draw the unseen facts of it. Machine Learning algorithms play a major role in the process of prediction[11]. There are broadly two categories of machine learning often called Supervised and unsupervised learning, where supervised learning makes out analysis on the gathered sets of data called experience, where we try to sort out the possible occurrences of the current problem looking into the past history of how to handle such similar issues. These can be done using algorithms that support Regression and Classification. The unsupervised category unlike prior one doesn't have any history and we need to derive a structure from the given data where Clustering is used. For Big data it is often preferable to do Feature selection, Classification and clustering.

A. Feature Selection: It is referred as a pre-processing step before data mining where we select a subset of features and remove the irrelevant features in order to reduce the computational complexity of the algorithm following approaches like Feature search done using any of the methods like complete, selection, Sequential and random search approaches. Feature evaluation describes a class of labels to describe the relevancy and correlation of one class of variables with another.[1][5] The feature evaluation uses few approaches like Wrapper, filter and hybrid approaches.

B. Classification: It is a supervised machine learning approach. There are several traditional classification mechanisms used to determine the data sets that fall under a particular class. A few approaches are like Decision tree (DT) as the name itself specifies it is a tree like structure where the data is organised at different levels. During retrieval process this structure requires a comparatively lower computational time, Support Vector Machine(SVM) This technique classifies the data based on the support

application in healthcare domain helps in diagnosis, prognosis, early intervention and optimal management of the data. There are few trade-offs that are to be handled as part of the future enhancements by the upcoming big data scientist like the technical support with respect to several platforms and patterns of Hadoop architecture, developing Skills and subject Professionals to handle several real time issues[5]. To provide an effective and secured transmission privacy is a major area to be focused[14][15].The visualization strategy also contains major issues like enhancing perceptual and interactive visualization capability. Also concentrate on meeting the need for speed, understanding data, Addressing data quality and displaying meaningful results.

References:

[1] Big data Analytics in Healthcare: A Survey Approach , Dharavath Ramesh1, Member, IEEE , Pranshu Suraj2, and Lokendra Saini3 Department of Computer Science and Engineering Indian School of Mines,
Dhanbad,Jharkhand-
26004,India,Email:ramesh.d.in@ieee.org1,{pranshusuraj2,lokendras903}@gmail.com

[2] Big Data in healthcare:Challenges and Opportunities, by *Hiba Asri OSER research team, FSTG Cadi Ayyad University Marrakesh, Morocco hiba.asri@gmail.com, Hajar Mousannif LISI Laboratory, FSSM Cadi Ayyad University Marrakesh, Morocco mousannif@uca.ma, 978-1-4673-8149-9/15/$31.00 ©2015 IEEE

[3] Predictive Big Data Analytics in Healthcare, A.Rishika Reddy Computer Science and Engineering Kakatiya Institute of Technology & Science Warangal, India rishika51896@hotmail.com, P. Suresh Kumar Computer Science and Engineering Kakatiya Institute of Technology and Science Warangal, India peddojusuresh@gmail.com, 2016 Second International Conference on Computational Intelligence & Communication Technology, 978-1-5090-0210-8/16 $31.00 © 2016 IEEE DOI 10.1109/CICT.2016.129

[4] Big Data: issues, tools and challenges by Avita Katal, Mohd Wazid, R H Goudar at IEEE 2013. 978-1- 4799-0192-0/13/$31.00 ©2013 IEEE

[5] Stephen Kaisler, Frank Armour, J. Alberto Espinosa, William Money, “Big Data: Issues and Challenges Moving Forward”, IEEE, 46th Hawaii International Conference on System Sciences, 2013.

[6] Big data analytics in healthcare: promise and potential Raghupathi and Raghupathi Health Information Science and Systems 2014, 2:3 http://www.hissjournal.com/content/2/1/3

[7] Big Data for Health Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, and Guang-Zhong Yang, Fellow, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19, NO. 4, JULY 2015

[8] Big data analytics ,FOURTH QUaRTeR 2011,By Philip Russom.

[9] Design Principles for Effective Knowledge Discovery from Big Data, Edmon Begoli, James Horey, 2012 Joint Working Conference on Software Architecture & 6th European Conference on Software Architecture

[10] Trends in bigdata analytics Karthik Kambatlaa, □,Giorgos Kolli asb ,Vipin Kumar c, Ananth Grama a, E-

mailaddresses:kkambatl@cs.purdue.edu(K.Kambatla),gkollias@us.ibm.com (G.Kollias),kumar@cs.umn.edu(V.Kumar),ayg@cs.purdue.edu(A.Grama),J.Parallel
Distrib.Computing journal homepage: www.elsevier.com/locate/jpdc.

[11] Sachchidanand Singh, Nirmala Singh, "Big Data Analytics", IEEE, International Conference on Communication, Information & Computing Technology (ICCICT), Oct. 19-20, 2012.

[12] M. Cottle, W. Hoover, S. Kanwal, M. Kohn, T. Strome, and N. W. Treister, Transforming Health Care Through Big Data, Institute for Health Technology Transformation, Washington DC, USA, 2013.