Survey: Importance and Research Issues in Big Data

Dinesh R
Project Manager, Shiro Software Solutions, Nagercoil, Tamil Nadu, India.

Dr.K.Selva kumar
Assistant Professor, Department of Mathematics, University College of Engineering, Nagercoil, Tamil Nadu, India.

Manibharathi R
Senior Developer, Shiro Software Solutions, Nagercoil, Tamil Nadu, India.

Abstract – In current time, the grow development of mobiles internet usage, Social Media, Cloud Computing, Internet of Things cause to the rapid increase of data in business, government, health care, and research in almost every discipline such as engineering, medical sciences, biological sciences, art & humanities. The basic aim of this paper is to explain the potential impact of big data importance, challenges facing in big data, open research issues and future research directions.

Index Terms – Big data, Research Issues, Significance, Research Issues.

1. INTRODUCTION

In current digital era, data are data collected from years and years of various source such as Machine, Human, Organization and the fast transition from digital technologies has led to growth of big data. In general, it refers to the collection of large and complex datasets which are difficult to process using traditional database management system tools or data processing system. These are available in unstructured, semi-structured, structured and format in petabytes and more. Formally, it is defined from 4Vs to 5Vs.4Vs refers to volume, velocity, and variety. Volume refers to the huge amount of data that are being generated everyday whereas velocity is the rate of growth and how fast the data are gathered for being analysis. Variety provides information about the types of data such as structured, unstructured, semi structured etc. The fourth V refers to veracity that includes availability and accountability. The Fifth V refers to Variability that contents Inconsistency of the data set can hamper processes to handle and manage it. The prime objective of big data analysis is to process data of high volume, velocity, variety, veracity and variability using various traditional and computational intelligent techniques [1]. The following Figure 1 refers to the characteristic of big data.

It is expected that the growth of big data is estimated to reach 66.79 Billion by 2021. Big data boost the software industries to the next level which will reach the $16 billion by 2025. Generally, Data warehouses are wont to manage the big dataset. During this case extracting the precise information from the on the market massive information may be a foremost issue. Most of the given approaches in data processing don't seem to be typically ready to handle the big datasets with success. The key drawback within the analysis of huge information is that the lack of coordination between information systems yet like analysis tools like data processing and applied mathematics analysis. These challenges usually arise once we would like to perform information discovery and illustration for its sensible applications. A basic drawback is a way to quantitatively describe the essential characteristics of huge information. There's a requirement for philosophy implications in describing information revolution. Additionally, the study on complexness theory big data can facilitate perceive essential characteristics and formation of advanced patterns in big information, modify its illustration, gets higher data abstraction, and guide the look of computing models and algorithms on big data.

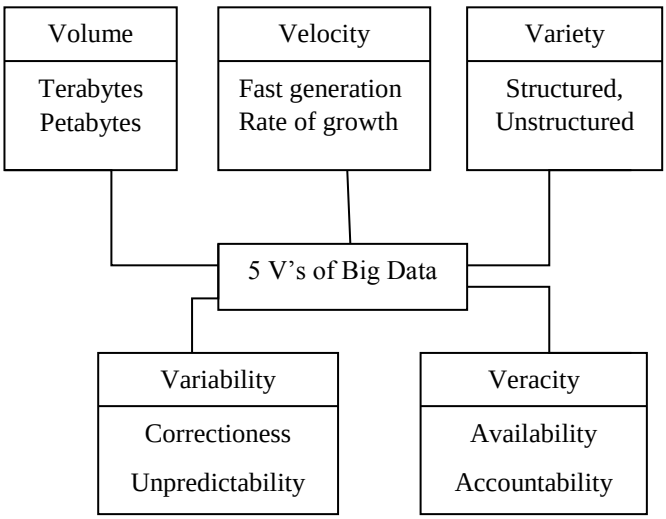

Fig. 1: Characteristics of Big Data

However, it is to be said that all data stored in the form of big data are tool useful for analysis and processing. Researcher is interested in disseminating the findings of the big data. This

work deals on challenges in big data and its available techniques. In addition to that we study the open research issues in big data.

2. IMPORTANCE OF BIG DATA

Big data transforms the way we live, work, and think due to its large volume of data. Here we describe the importance of big data in various points of views.

2.1 Importance of Big data in Government

In modern digital era, big data can have an enormous import on local, national and global. By using a big data analysis, governments can access vast amounts of relevant information to their daily functions. Big data is important because it not only allows the government to pinpoint areas that require attention, but it also gives them that information in real time. it allows governments to make faster decisions and t it allows them to monitor those decisions and enact changes if necessary. Here we describe the following areas that big data can positively affect at the government level are transportation, health care, education and poverty [3].

With Big Data, authorities will generate a lot of precise understanding of the client demand on totally different routes. They will map client journeys across multiple modes of transportation – trains, buses, non-public modes of transportation etc. With Big Data, information from the sensors put in on the instrumentation will be analyzed at a lot of quicker rate and at a lot of minute-level. This may be wont to predict coming faults at the individual element levels like brakes, a stretch of rails etc. With this, authorities will schedule maintenance of the instrumentation exactly at the correct time. With Big Data, authorities will have associate intimate understanding of client journeys – begin and finish points involving trains, buses and even non-public modes of transportations. It is important to own quick and best response particularly throughout unplanned events and massive information technologies nowadays give tools to realize like response.

Government can minimize unemployment rate by predicting the task desires before primarily based the attainment rate. This will be achieved by analysis the scholars graduating annually. It allows government to rearrange for special trainings so as to create young entrepreneurs [5]. Education is one in all the dear assets which will lean to the youngsters. It's the duty of state to produce quality education to children [4]. Big data analytics provides elaborate report of youngsters World Health Organization are within the age to be admitted to the varsity. This helps government to assess the academic desires for these kids forthwith. Today individuals got to sit up for a protracted time to urge EB, telephone, water, identity card and gas association. These are the essential desires of national. It's the responsibility of the govt. to produce these services as fast as possible. Big data analytics plays a serious role in achieving it as a result of the information is analyzed on daily. People that are in would like are served forthwith.

2.2 Importance of Big data for Health Care

By exploitation Big data patient central services, we will offer quicker relief to the patients by providing proof based mostly medicine-- detective work diseases at the sooner stages supported the clinical knowledge offered, minimizing drug doses to avoid aspect impact and providing economical medication supported genetic makeups[6]. This helps in readmission rates thereby reducing price for the patients. By detective work spreading diseases earlier, we will predict the spreading diseases earlier before spreading supported the live analysis. This could be known by analyzing the social logs of the patients plagued by a unwellness during a specific geo-location [6]. This helps the attention professionals to advise the victims by taking necessary preventive measures.

By watching whether or not the hospitals area unit setup per the norms setup by Indian medical council. This periodical check-up helps government in taking necessary measures against disabling hospitals.

By exploitation huge knowledge analytics we will custom-made patient treatment that watching the impact of medication incessantly and supported the analysis dosages of medicines are often modified for quicker relief. Watching patient very important signs to produce proactive care to patients. Creating Associate in nursing analysis on the information generated by the patients United Nations agency already suffered from constant symptoms, helps doctor to produce effective medicines to new patients.

2.3 Importance of Big data for Business

In the current business situation, the progress and also the success of the business depends on the proper call taken among varied choices accessible. The standard and amount of underlying knowledge verify the proper call. Historically, MIS, DSS, Roman deity and EIS were used because the business deciding systems, that were knowledge-oriented and restricted in data capture, storage and analytics capabilities. But Big data is largely immense quantity of knowledge that cannot be effectively captured, processed and analyzed by ancient information and search tools in cheap quantity of your time.

With the introduction of conception of Big data, technologies like knowledge visualization, large multiprocessing architecture, cloud computing platforms, distributed file systems and scalable storage systems came to forefront, resulting in attention-grabbing applications from various areas like drugs, insurance and ecommerce. Big data analytics (BDA) has been given vital attention by businesses in recent years because of the promise of upper productivity and profit [7]. Big data will produce period of time solutions to

challenges in each industrial sector [9]. There are many samples of firms that are able to become huge knowledge drivers with success.

Big data offers several benefits for business, such as, improved deciding, presents new insights, and create processes additional economical, it's not nevertheless convinced most European organizations of manufacturing worthy returns from big data investments [8].

2.4 Importance of Big data for Scientific Research

Big data is transforming how scientists conduct research traditionally. The World Universities are now changing research technique with big data technologies. Grants and research programs are given at improving the core technologies around managing and working big data sets, and speeding up scientific research with big data. The emergent field of data science is changing the direction and speed of scientific research by letting people fine-tune their inquiries by mentoring into giant data sets.

According to Huadong Guo, Editor-in-Chief, International Journal of Digital Earth, emphasizes the need to develop a better understanding of the role of big data for scientific research, thereby strengthening international science for the benefit of society by developing policy frameworks, research guidelines, case studies, and best practices which will help us exploit big data effectively [10].

3. CHALLENGES IN BIG DATA ANALYTICS

Recent years big data has been acquired in several domains like health care, government, education, retail, business and other interdisciplinary scientific researches. Web-based applications encounter big data frequently, such as power prediction, user behavior, social media computing, healthcare, content recommendation systems, internet text and documents, internet search indexing and the smart city. Taking into account these advantages of big data it gives new opportunities in the knowledge processing applications for the further researchers. However opportunities always follow some challenges.

3.1 Storage Medium and Analysis

In recent days the size of data has grown exponentially by various means such as mobile devices, web, social media, aerial sensory technologies, remote sensing, bank sector, health care etc. These big volumes of data required to store analytic, compliance with laws and service level agreements to protect and preserve data. These data hold on outlay spending much cost whereas they unnoticed or deleted finally because there's no enough area to store them. Therefore, the primary challenge for big data analysis is storage mediums and better input or output speed.

In spite of the large advantages that can be gained in big data set, large data demand for storage and processing poses a major challenge. Big data just keeps getting bigger. According to IDC, by 2020 the digital universe will expand from 4.4 zettabytes to 44 zettabytes. Technology trends like mobility and the internet of things (IoT) are generating tons of data that can fuel crucial business decisions [11]. Existing traditional techniques cannot support and perform effective analysis, due to the big data volume, variety, and velocity of data.

According to Jianqing Fan et al , Big data are characterized by high spatiality and enormous sample size. These two options raise three distinctive challenges: (i) high spatiality brings noise accumulation, spurious correlations and incidental homogeneity; (ii) high spatiality combined with big sample size creates problems like significant procedure value and algorithmic instability; (iii) the huge samples in massive knowledge square measure generally mass from multiple sources at completely different completely different} time points victimization different technologies. This creates problems with heterogeneousness, experimental variations and applied math biases, and needs U.S. to develop additional adaptative and strong procedures [12].

Due to the size of big data keeps increasing exponentially, the available traditional tools cannot be efficient to process these data for obtaining meaningful information. The unique characteristic of Big Data is the process in its value is discovered knowledge. Big Data is unlike conventional business intelligence, where the simple reporting of a known value reveals a fact, such as summing daily sales into year-to-date sales. With Big Data, the goal is to be clever enough to discover patterns, model hypothesis, and test your predictions. Subject matter experts interpreting visualizations or making interactive knowledge-based queries can be aided by developing machine learning adaptive algorithms that can further discover meaning [13].

3.2 Computational Complexities

The computation of big data becomes difficult with traditional computing technologies such as data mining, machine learning due to large volume, velocity, variety, veracity and variability of data. The big data computations are not rely on old statistics, analysis tools and iterative algorithms used in small size of data in traditional perspectives. New approaches will require to solving problems involving in big data. We have to rethink and explore its computability, computational complexity and algorithms [14].

3.3 Scalability and Visualization of Data

With the tremendous growth of dataset sizes, the scalability of most current feature choice algorithms may well be became risk. In several scientific and business applications, information is typically measured in terabytes. Normally,

datasets within the scale of terabytes can't be loaded into memory directly, thereby limiting the usability of most feature choice algorithms [14]. Currently, there square measure some makes an attempt to use distributed programming frameworks like Map Reduce and MPI to perform parallel feature choice for terribly large-scale datasets. Because the information size is scaling abundant quicker than central processing unit speeds, there's a natural dramatic shift in processor technology being embedded with increasing variety of cores [15]. This shift in processors leads to the event of parallel computing. Real time applications like navigation, social networks, finance, web search, timeliness etc. needs parallel computing.

Data visualization required to represent knowledge more effectively by using different graphical representation. To carry information easily by showing knowledge hidden in the complex a big data sets, attractive form and functionality are necessary. Information that has been abstracted in some schematic forms, including attributes or variables for the units of information, is also valuable for data analysis. This way is much more instinctual [169] than innovatory approaches. The current Big Data visualization tools mostly have poor performances in functionalities, scalability and response time. Now we have to rethinking the way we visualize Big Data, not like the way we adopt before.

3.4 Information Security

Big Data transforms High Volume, Velocity and Variety of data into Value. Every company has their own policy to protect privacy information of user. Big data have big security threat related protecting sensitive information. According to researcher, the big data has the following list of security challenges. They are Protected database storage and transaction log file, Secure computations in distributed frameworks, Privacy issues for non-relational data stores, End-point input validation/filtering, Extensible and pasture able privacy preserving data mining and analysis, Real-time security and compliance monitoring, Cryptographically enforced access control and secure communication, Granulated access control, Information security, Information security [16].

4. OPEN RESEARCH ISSUES IN BIG DATA ANALYTICS

Big data analytics involves the processes of searching a database, mining, visualization, transfer and analyzing data dedicated to improve company performance. The amount of data is increasing exponentially in big data but the data processing technologies are used improving very slowly. Numerous tools are available to address the key issues of big data analytics. The current data processing applications such as Map Reduce, Dryad, Pregel, PigLatin, MangoDB, Hbase, SimpleDB, and Cassandra cannot solve the real problems of storing and querying big data.

4.1 IoT for Big Data Analytics

Big data analytics is quickly rising as a key IoT initiative to boost higher cognitive process. Knowledge acquisition from IoT information is that the biggest challenge that huge information skilled face. Therefore, it is essential to develop infrastructure to investigate the IoT data. An IoT device generates continuous streams of knowledge and therefore the researchers can develop tools to extract substantive data from these information victimization machine learning techniques. Understanding these streams of knowledge generated from IoT devices and analyzing them to induce substantive data could be a difficult issue and it ends up in big data analytics. Machine learning algorithms and machine intelligence techniques is that the only answer to handle huge information from IoT prospective.

In the context of massive IoT data, security and privacy ar the key challenges in process and storing large amounts of data. Moreover, to perform vital operations and host non-public data, these systems extremely place confidence in third party services and infrastructure. Therefore, AN exponential growth in rate causes problem in securing every and each portion of vital data. As antecedently mentioned, existing security solutions aren't any long applicable to providing complete security in huge IoT knowledge situations.

With relation to data generated through IoT, Dethlefs et al. [17] the following security issues: timely updates - problem to keep systems up to this point, incident management - distinctive suspicious traffic patterns among legitimate ones and doable failure to capture unidentifiable incidents,) ability - proprietary and vendor-specific procedures can cause difficulties find hidden or zero day attacks, and protocol convergence - though IPv6 is presently compatible with the newest specifications, this protocol has nonetheless to be totally deployed.

Therefore, the appliance of security rules over IPv4 might not be applicable to protective IPv6.Moreover, synchronization problems could occur in parallel computing, whereas info is changed inside totally different data processing strategies. This bottleneck data mining strategies has become Associate in Nursing open issue in big IoT data analytics that ought to be self-addressed. Coming up with visualization answer that's compatible with advanced huge knowledge categorization frameworks could be a troublesome task.

Similarly, time interval could be a fascinating think about data IoT data analytics. Consequently, cloud computing architectures supported with wealthy graphical user interface facilities are often deployed to get higher insights into data IoT data trends.

4.2 Cloud Computing for Big Data Analytics

In order to move beyond the existing techniques and strategies used for machine learning and data analytics, some challenges need to be overcome. NESSI [18] identifies the following requirements as critical. There are: In order to select an adequate method or design, a solid scientific foundation needs to be developed. New efficient and scalable algorithms need to be developed, for proper implementation of devised solutions; appropriate development skills and technological platforms must be identified and developed. Lastly, the business value of the solutions must be explored just as much as the data structure and its usability.

Talia [19] suggests further research and development in the following areas: Programming abstracts or scalable high-level models and tools, Solutions for data and computing interoperability issues, Integration of different big data analytics frameworks, Techniques for mining provenance data.

Some threats and issues, such as confidentiality, privacy, integrity, and availability of data, exist in big data using cloud computing environment. Therefore, data security must be considered once data are outsourced to cloud service providers. The cloud must also be evaluated at regular intervals to protect it against threats. Cloud vendors must make sure that all service level agreements are met. Recently, some dissensions have leaked how some security agencies use data generated by individuals for their own benefit without permission. Therefore, policies that cover all user privacy concerns should be developed. Traditionally, the most common technique for privacy and data control is to protect the systems utilized to manage data rather than the data itself; however, such systems have proven to be vulnerable. Utilizing strong cryptography to encapsulate sensitive data in a cloud computing surroundings and developing a novel algorithm that efficiently allows for key management and secure key exchange are important to manage access to big data, particularly as they exist in the cloud independent of any platform. Moreover, the issue with integrity is that formerly elaborated hashing schemes are no longer applicable to huge amounts of data. Integrity verification is also hard because of the lack of support, given remote data access and the lack of information on internal storage [20].

4.3 Visual Computing & Multimedia Analytics

The large-scale multimedia data generated daily life has opened opportunities and challenges for intelligent multimedia analysis. Meanwhile, with the recent advances in deep learning techniques, we are now able to boost the intelligence of multimedia analysis significantly and initiate new research directions to analyze multimedia content. Therefore, deep learning for intelligent multimedia analysis is becoming an emerging research area in the field of multimedia and computer vision. The following research direction have to go 1) Multimedia Retrieval (image search, video search, speech/audio search, music search, retrieval models, learning to rank, hashing) 2) Web IR and Social Media (link analysis, click models, user behavioral mining, social tagging, social network analysis, community-based QA) , 3) Deep image/video understanding (object detection and recognition, localization, summarization 4) highlight detection, action recognition, multimedia event detection and recounting, semantic segmentation, tracking), 5) Vision and language (image/video captioning, visual Q&A, image/video commenting, storytelling), 6) Multimedia data browsing, visualization, clustering and knowledge discovery 7) Home/public video surveillance analysis (motion detection and classification, scene etc.

4.4 Spatial & Context-Aware Data Management

Spatial & Context-Aware Data Management have to addresses core topics on the design, the use and the evaluation of Big Data enabling technologies to build next-generation context-aware applications and computing systems for future intelligent environments. Disruptive paradigm shifts such as the Internet of Things (IoT) and Cyber-Physical Systems (CPS) will create a wealth of streaming context information. Large-scale context-awareness combining IoT and Big Data will drive to creation of smarter application ecosystems in diverse vertical domains, including smart health, finance, smart grids and cities, transportation, Industry 4.0, etc. However, effectively tapping into growing amounts of disparate contextual information streams remains a challenge, especially for large-scale application and service providers that need timely and relevant information to support adequate decision-making. A deeper understanding is necessary on the strengths and weaknesses of state-of-the-art big data processing and analytics systems (Hadoop, Spark, Storm, Samza, Flume, Kafka, Kudu, etc.) to realize large-scale context-awareness and build Big Context architectures. In particular, the key question is how one can help identify relevant context information, ascertain the quality of the context information, extract semantic meaning from heterogeneous distributed information sources, and do this data processing effectively for many concurrent context-aware applications with different requirements for adequate decision-making.

The following area are hot research area for Spatial & Context-Aware Data Management Big Data architectures for large-scale context-aware applications, Context models and query languages for heterogeneous data streams, Distributed context reasoning with Big Data technologies, Machine learning and prediction of situational awareness with Big Data, Effective data collection and processing for concurrent context-aware applications, Modeling of Quality of Service constraints and enforcing of Service Level Agreements,

Context-aware dynamic decision making on streaming Big Data Context-driven monitoring, adaptation and optimization of Big Data systems Large-scale Quality of Context management Systematic comparison of Big Data technologies for context-aware applications Big Context solutions for finance, health, smart cities, industry 4.0, Security, privacy, scalability, and sustainability concerns Big Context systems.

4.5 Recommender Systems & Preference Analytics

Recommender systems research has made significant advances over the past decades and has seen wide adoption in electronic commerce. Recently, a variety of types of side information such as social friends, item content has been incorporated into recommender systems to further enhance their performance, especially the well-recognized problem of data sparsity. However, most of existing approaches have only investigated the value of a single type of side information at a time, such as social trust, friendship, or item contents. In real-life applications, users may have different kinds of reactions towards items of interest.

Recommender Systems & Preference Analytics research focus on technologies and solutions related to the following topics such as Algorithms that exploit homogenous and heterogeneous side information to make better top-N item recommendations, and that address data sparsity and cold start issues. Systems that can apply side information to enhance recommendation diversity, novelty and explain ability, Context-aware recommendation systems incorporating side information, Large-scale parallelization techniques to speed up information fusion and generation of personalized recommendations, Incremental recommendation solutions to handle continuous updates, especially real-time streaming data for recommendations. Innovative, efficient recommendation tools and benchmarks to integrate multiple types of side information and enhance reproducibility and comparison of such recommendation models.

4.6 Natural Language Processing & Text Mining

The Natural Language Processing & Text Mining fields are facing many new challenges arising from intelligent applications and big data, such as business intelligence, social analytics, etc. The Seventh Conference of NLPCC will focus on Natural Language Analysis and Understanding in Big Data including Fundamental Research in Language Computing, Multilingual Access, Web Mining/Text Mining, and Machine Learning for NLP, Knowledge Graph, Information Extraction, Question Answering, Information Retrieval, Chat bot, NLP for Social Networks, as well as the Applications of Language Computing.

The Natural Language Processing & Text Mining research is focus on the following related topics Word Segmentation & Named Entity Recognition, Syntactic Parsing, Semantic Analysis, Discourse Analysis, NLP for Minority and Low-resource Languages, Applications of Natural Language Processing, Digital Publishing, Document Engineering, OCR & Fonts Computing, NLP for Mobile Computing, Machine Translation & Multilingual Information Access, Machine Learning for NLP, Web/Text Mining & Big Data, Information Retrieval & Extraction, Knowledge Representation & Acquisition, Personalization & Recommendation, Semantic Web & Knowledge Base, NLP for Search & Ads, Query & Document Understanding, Topic Detection & Tracking, Sentiment analysis & Opinion Mining, Question Answering, Chatbot & User Interaction, Social Computing, Rumour Detection.

4.6 Crowd sourcing and Pervasive Computing

With the rapid development of metropolitan areas, urban problems, such as environment pollution, traffic congestion and increased energy consumption, increase more and more. Smart cities are emerging as a novel and promising paradigm to tackle existing problems and avoid potential issues in our cities. Recently, with the pervasive usage of intelligent equipment – e.g., GPS devices, GIS tools, traffic cameras, smart cards, smart phones and road deceleration devices , Mobile Crowd Sourcing (MCS) enables citizens to collect and share data to make our cities smart.

MCS can be applied in several urban scenarios, including traffic planning, mobile social recommendation, public safety, and so on. Meanwhile, pervasive computing has been widely used to develop smart cities by utilizing emerging computing technologies. However, several challenges still exist for an effective use of MCS and pervasive computing technologies in smart cities. Research challenges include MCS architecture to sense urban pulse and investigate the spatio-temporal mobility patterns comprehensively, compute and analyze urban big data generated by large amount of participants/devices, multi-sourced and heterogeneous data to facilitate applications towards smart cities.

To address these challenges and find effective solutions for smart cities, this special issue focuses on mobile crowd sourcing and pervasive computing research challenges in urban scenarios. Specifically, researchers are encouraged to to focus the following area Mobile crowd sourcing technology for smart cities Crowd sourcing systems and services for smart cities Application of pervasive and mobile computing technology for smart cities Fault tolerance, reliability and survivability in smart cities Vehicular social networks in smart cities Network communications technology for smart cities Data mining and analytics of urban big data Safety, security, privacy and trust in applications and services for smart cities Human mobility models for smart cities Multi-sourced heterogeneous data acquisition and fusion for smart cities Case studies and test beds.

4.7 Big data for Health Care

Big data provides the following advantages to healthcare industries such as improve quality of care, managing population health, early detection of diseases, data quality, structure, and accessibility, improve decision making, cost reduction, patient-centric care, enhances personalized medicine, globalization, fraud detection, and health-threat detection [21].

Applying computation model to huge-scale datasets is the primary goal for health big data Analytics. The Complexity of the analysis in which the computing time for finding solutions increases exponentially as the number of records increases, scale of the data in which computing performance will decrease dramatically when data scale and diversity increase, Parallelization of computing model in which Many statistical analysis approaches or data mining algorithms are difficult to parallelize, Availability of computing resources a significant cost associated with acquiring and maintaining the Super computing resources for solving computationally intense problems are the key open issues in healthcare big data analytics.

Big data applications in genomics cover a wide variety of topics. Here we focus on pathway analysis, in which functional effects of genes differentially expressed in an experiment or gene set of particular interest are analyzed, and the reconstruction of networks, where the signals measured using high-throughput techniques are analyzed to reconstruct underlying regulatory networks. These networks influence numerous cellular processes which affect the physiological state of a human being [22].

A large amount of data is produced in intensive care units (ICU) from each patient. However, the potential for developing clinical decision support systems (CDSS) in an ICU environment has given new opportunities to researchers. The patient care management system has been proposed which combines traditional data and stream data recorded from critically ill patients in the ICU for data mining and alerting medical staff of critical events in real time. However, this field is still at an early phase of its development.

5. CONCLUSION

We survey the importance of big data, challenges of big data analytics, and open research issues related to big data and it environment used to analyze these big data. From this survey, it is understood that every big data platform interrelated to each other. But advancement in big data reflects all type of industries and current technologies. It produces new opportunities as well as new challenges all the area of research. We have to address the issues related to big data area such as IoT for Big Data Analytics, Cloud Computing for Big Data Analytics, Visual Computing & Multimedia Analytics, Spatial & Context-Aware Data Management, Recommender Systems & Preference Analytics, Natural Language Processing & Text Mining and Crowd sourcing and Pervasive Computing to solve problems of big data effectively and efficiently.

REFERENCES

[1] M. K.Kakhani, S. Kakhani and S. R.Biradar, Research issues in big data analytics, International Journal of Application or Innovation in Engineering & Management, 2(8) (2015), pp.228-232.

[2] https://www.marketsandmarkets.com/PressReleases/big-data.asp

[3] https://datafloq.com/read/5-applications-big-data-in-government/65

[4] Jean Yean, Big Data Bigger Opportunities, White Paper, U.S.General Services Administration.

[5] Bill Hamilton, Big Data Is the Future of Healthcare, Cognizant white paper, 2010.

[6] Yanglin Ren, Monitoring patients via a secure and mobile healthcare system, IEEE Symposium on wireless communication,2011.

[7] S. Akter, S. Fosso Wamba, A. Gunasekaran, R. Dubey, and S. J. Childee, "How to improve firm performance using big data analytics capability and business strategy alignment?," Int. J. Prod. Econ., vol. 182, pp. 113–131,2016.

[8] F. Portela, L. Lima, and M. F. Santos, "Why Big Data? Towards a Project Assessment Framework," Procedia Comput. Sci., vol. 98, pp. 604–609, 2016.

[9] D. Sen, M. Ozturk, and O. Vayvay, "An Overview of Big Data for Growth in SMEs," Procedia - Soc. Behav. Sci., vol. 235, pp. 159–167, 2016.

[10] Huadong Guo,Big data for scientific research and discovery,Feb 2015.

[11] http://www.vicomnet.com/top-data-storage-challenges/

[12] Jianqing Fan , Fang Han and Han Liu,Challenges of Big Data analysis, National Science Review, Volume 1, Issue 2, 1 June 2014.

[13] An Enterprise Architect's Guide to Big Data Reference Architecture Overview, Oracle Enterprise Architecture White Paper, March 2016.

[14] S. Singh et al., "Parallel Large Scale Feature Selection for Logistic Regression," Proc. 2009 SIAM Int'l Conf. Data Mining, 2009, pp. 1172–1183.

[15] H. Zhu, Z. Xu and Y. Huang, Research on the security technology of big data information, International Conference on Information Technology and Management Innovation, 2015, pp.1041-1044.

[16] Aditya Dev Mishra, Youddha Beer Singh,Big Data Analytics for Security and Privacy Challenges,2016.

[17] Dethlefs, T., et al. Energy Service Description for Capabilities of Distributed Energy Resources. in DA-CH Conference on Energy Informatics. 2015. Springer.

[18] NESSI, Big Data: A New World of Opportunities ,2012.

[19] Talia, D. ,Clouds for Scalable Big Data Analytics. Published by IEEE Computer Society. 2013.

[20] brahim AbakerTargioHashem,n, IbrarYaqoob , NorBadrulAnuar , Salimah Mokhtar, AbdullahGani , SameeUllahKhan, The riseof "big data" on cloudcomputing:Reviewandopen researchissues , Information Systems47(2015)98–115.

[21] Clemens Scott Kruse, Rishi Goswamy an Sarah Marawi, MHA,Challenges and Opportunities of Big Data in Health Care: A Systematic Review, JMIR Med Inform. 2016 Oct-Dec; 4(4): e38.

[22] Ashwin Belle,Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian, Big Data Analytics in Healthcare,Volume 2015,Article ID 370194.

Authors

Mr. Dinesh R has obtained his Bachelor Degree in Physics from Manonmaniam Sundaranar University. The obtained his Masters Degree from Anna University. Currently He is working as Project Manager in Shiro Software Solutions. He has 6 years experience in Software industries. He also has 4 years experience in teaching as Assistant Professor. His Specializations include Cloud Computing, Big data and Networkig.

Dr. K. Selva kumar started my research and teaching works from April 1987 at Bharathidasan University, Trichy, Tamilnadu, India. Received Ph.D. from Bharathidasan University in 1992. At present working as Assistant Professor in Mathematics at University College of Engineering, Anna University, Nagercoil Campus, Tamilnadu, India. Developing software for Networking , cloud computing. Image Processing, Singular perturbation problems in control design, aircraft optimal control guidance, Numerical methods for engineering related research problems.

Mr. R. Manibharathi completed MCA from Institute Of Road and Transport Technology, Erode, Anna University, Chennai. He graduated BSc Information Technlogy from Manonmaniam Sundaranar University. He is currently working as Senior Developer at Shiro Software Solutions. His research interest in Cloud Computing, networking and Big Data.