

Big Data Security: Current State and Future Opportunities

Ritesh Kumar

riteshchandel@gmail.com

Ph.D. Scholar, Department of Computer Science and Engineering, U.I.E.T., M.D.U. Rohtak, Haryana

Dr. Anuj Kumar Sharma

anuj.k.er@gmail.com

Principal, BRCM College of Engineering & Technology, Bahal, Bhiwani, Haryana

ABSTRACT

An incredible advancement & trends are noticed in industries driven by data as the new oil, automation, methodologies and computation. Due to researches, industrial processes and human activities the data collection and processing has increased to an unprecedented scale. Many business players and the technology providers are currently working in the generation of award-winning products, services and also proposing trendy business models based on data classification, prediction, aggregation and qualitative analysis.

One of the main issues that arise by using the "Big Data" System is the privacy & security. The "Big Data" systems are tedious, difficult to manage and composite in nature hence the security must be thoroughly checked & assured. Also due to integration of multitude of latest tools & procedures, the new security limitations are introduced and need to be addressed in the true sense.

This report is an amalgamation & exploratory study of analyzed different inputs from different individualistic sources so that we better understand the "Big Data" analytics in different ways. This dissertation limelight's the use of the "Big Data" system by the private organizations in the sectors like Finance, Energy, and Telecom. With regard to the analysis outcome of theories & "Big Data" privacy issues, recommendations to the organizations are provided to support secure adoption of "Big Data" system.

Keywords: Big data, Data sharing, Data privacy, Security, Data Ownership.

INTRODUCTION

In the recent year the advent of latest digital tools & technologies has made it possible to allow people to make fruitful decisions, behavioral patterns and adjustments. The "Big Data" introduces the capabilities to throw some light on the perishing needs and evaluating them, too with regard to a broader and speedier manner. Hence, competitive advantages are provided to the businesses by using the "Big Data" which may identify the opportunities to execute the targeted changes in order to improve continuously.

"Big Data" is known an invaluable asset and, as just that, is designed to target by cyber threats. according to Enisa's Threat Landscape (Sam Madden,2012) [21]. In addition, it's predicted to get to be a very useful resource for security professionals, as it can make a major contribution to building intelligence, threat detection and incidental detection. The "Big Data" in this paper is defined and the main focus is given to the emerging information security challenges faced in "Big Data" in three different important sectors like Finance, Energy and Telecom. The mitigation measure and the recommendation are provided to avoid these challenges in the companies.

The data is constantly generated by people and devices and is being increased day by day over last 20 years and the web is overloaded with huge and exponential generation of data. A 2011 study by the “International Data Corporation (IDC)” revealed that the total worldwide data volume generated and replicated was 1.8ZB (almost 1021B), which increased by almost nine times in five years (Omer Tene et.al.) [14]. Actually, this figure was declared to almost double the near future at least every other 2 years.

Big Data

A large number of sensor devices & people are connected to a global network that manipulates the potential in generating, communicating, sharing and accessing the data. In the recent years the volume of data has increased at such an extent that it no more can be handled and processed by the traditional database system.

“Big Data” may be defined as the extremely huge amount of data that need to be analyzed for computation in order to draw a meaningful pattern or format out of it. There are many definitions presented in different articles. Both mentioned

previously communities of authors go far beyond the single dimensional aspects of data when defining “Big Data”! EddDumbill in (Samuel Fosso et.al.) [22] clearly conveys Big Data’s multidimensionality when specifying that “the data is too big, moves too fast, or doesn’t fit the structures of your database architectures”. The “Big Data” are characterized by its High volume, high variety, high speed, high accuracy and good value. The “Big Data” have the 5 v’s properties in them which are: Volume, Velocity, Variety, Veracity and Validity

Big Data Applicationary-area domains

(Divyakant Agrawal,) [4], Public as well as private sector companies has started following the “Big Data” solution in order to increase value in their business services and optimizing their internal process. We have provided Standard visualization of high-level application architecture intended per segment in fig.1.

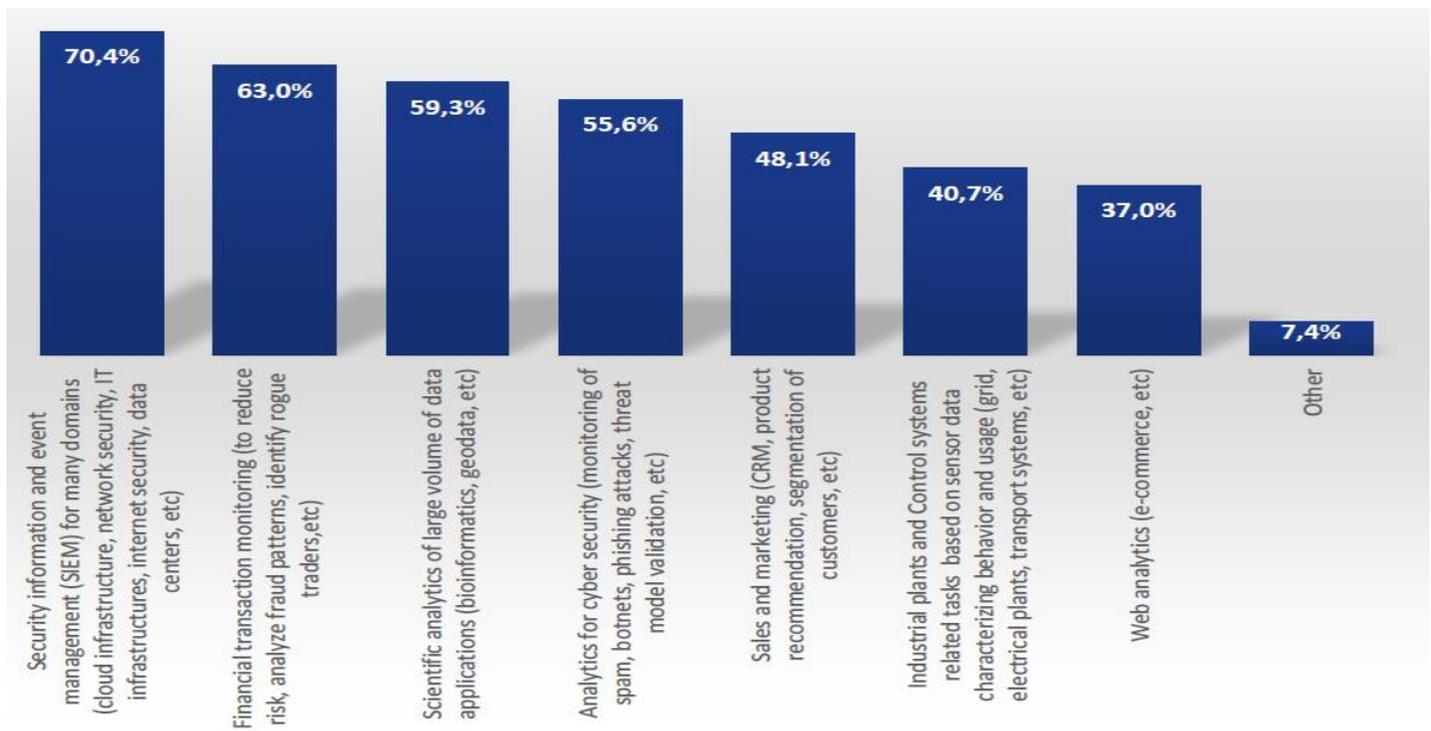


Fig 1: The far more important field of “Big Data” applications

Big Data Security

(Hongbo Zou, Yongen Yu) [8], The data is the most important asset for an enterprise. Many enterprises and businesses are using data for marketing and research so must be sure from the security perspective. Not only the trending word for businesses the “Big Data” is more than that, it provide the fruitful benefits to those who uses it willingly. Since the “Big Data” are used by many companies it doesn't mean that it has made its road to the future. There are many risk, issues, security breaches and challenges present along with it which are under “Big Data” security. There are many organizations that uses “Big Data” for storing and analyzing petabyte of data coming from web logs, social media and click stream to acquire knowledge about their business and customer.

(S. Garg and M. V. A.K.R) [29], The security breach may lead to more serious and damaging. Hence the main concern to the Information Technology is the privacy and safety issue. The challenges in the safety issue is because of data encryption to the heterogeneous data collected, and there are two aspects in the “Big Data” privacy challenges i.e. one during acquisition of data second during the storage, transmission and usage.

REVIEW OF LITERATURE

Presently a-days Data is getting increasingly important yet how to handle data and finding concealed realities from it is increasingly significant. “Big data” is a wide term for any voluminous and compound datasets for example excessively huge, quickly developing, and hard to handle by utilizing traditional apparatuses and strategies (Mathur R.) [12].

“Big data” can be created by means of different sources like cell phones, sensors, sound and video sources of info and online networking, are on the whole expanding the volume and assortment of data (Rohit Chandrashekar, et. Al)[19]. “Big data” can possibly give important data in the wake of preparing that can be found through profound investigation and effective preparing of data by decision makers. To remove significant experiences from such fluctuated and quick developing datasets, different instruments and strategies of “Big data” Analytics can

be utilized that may prompt better decision making and strategic planning.

(V Srinivas, P Srikanth)[25], in her exploration on “Big data”, reasons that the 3V's of “Big data” will develop continually in intricacy. The data will show up quicker in substantially more perplexing structure and in progressively higher volumes. Much of this data will be valuable, at times only couple of moments or even less.

As indicated by (S.M. Drus and N.H. Hussan) [38], "The volume of data worked upon by present day applications is developing at a gigantic rate so it has made the charming difficulties for equal and appropriated processing stages". These challenges are risen up out of building stockpiling frameworks that can oblige these huge datasets to gathering data that can procured from endlessly topographically conveyed sources.

As per (Dr Saravana kumar N M et.al)[5], "Big data advances are valuable in human services industry. By utilizing it we can change the entire human services esteem fasten from tranquilize examination to patients caring quality. In any case, the unstructured nature of “Big data” of wellbeing industry is found. Thus, it is important to structure the Big data of wellbeing industry"

(Srivathsan, M and Yogesh, Arjun K, et. el)[23] additionally depict mechanical progression in human services area. This progression is conceivable just by the execution of prognostic registering incorporates the Big data investigation. In this procedure organized and unstructured biomedical data can be gotten from a wide range of analyses and reviews gathered by medical clinics, research centers, pharmaceutical organizations or even web based life. By this forecast registering the productivity what's more, expectation for everyday comforts of individuals can be created.

As per (Raghunath Nambiar et. Al)[18], with the assistance of “Big data”, more precise medications can be utilized by the patient based on explicit data for example, genomics and proteomics. This data can be made on the profiling of comparable patients.

(Ping Jiang and Jonathan Winkley et.al)[15], examined and broke down the data created by the wearable sensors. They introduced a "Big data” medicinal services framework for older

individuals". Such a "Big data" framework can give rich data to human services suppliers about people's wellbeing conditions and their living condition. In this manner showed the need of the "Big data" innovation in gathering what's more, dealing with the data created.

(Ruchie Bhardwaj et.al)[20], has talked about the "Big data" innovation in medicinal services industry. As per the specialist, the five worth pathways are comprised of right living, right consideration, right suppliers, right worth and right development. They characterize the system of the new business. These methodologies lead to a more fruitful treatment for patients. What's to come is brilliant for the most up to date convergence among innovation and human services.

(Arthur R. Miller, Charles R. Ashman) [2], stress in their investigation the significance of "Big data" for wellbeing arranging framework. In this innovation abundant patient data and chronicled data are put away for the investigation.

(Marco Viceconti, Peter Hunter, and Rod Hose;e.et)[11], likewise put accentuation on utilizing the "Big data" advancements to investigate the data put away and bringing better bits of knowledge. This will limit the danger of examination speculations, and will guarantee a steady improvement of in silico medication, preferring its clinical reception.

(Tetsu Akiyama) [24], advocated a further method for a single hidden deep neural network incorporating "ELM NMF (non-negative lattice factorization)" as well as Outrageous Learning Machine (OLM). Both strategies are consistent with exceptionally large datasets and thus are efficient and powerful technologies.

(R Madhuri et.al), [17], found three new techniques used in the clustering of large data sets. This same method used was gradual "K-means" where groups were using math information. In the strengthened calculation of k-modes, the bunches used positive data; while the k-model was positive on the basis of the mixed number juggling and the distinct figures.

(Ganesh Krishnasamy et.al.) [6], anticipated the improved Cohort Insight (MCI) methodology in his research by assigning activities such as shifting the "worker to Cohort Intelligence (CI)" to overpower the worry of predominance and combination

quickness. In addition, the creators proposed a combination of different algorithms, such as the KMCI strategy for the massive data segmentation, which would be a conglomeration including its conventional K-Means and the MCI and has the advantage of both algorithms.

(Hong Yu et al.) [7] foreseen a capable automatic approach to handle the issue of deciding the measure of clusters. After this the creators have advertised "ACA-DTRS (programmed grouping algorithm dependent on DTRS)" a procedure in light of various leveled grouping, which routinely quits bunching once finding the necessary measure of clusters. They additionally recommended "FACA-DTRS" which is a snappier sort of "ACA-DTRS" as far as entanglement. The two methods are capable as far as time along with cost.

(Buza et.al.) [10], introduced an approach to concentrate that is squandered even in data of the tick is growing rapidly. Utilizing a novel grouping algorithm called SOHAC the fundamental tick data is isolated by utilizing grouping highlights into littler "data matrix". They additionally introduced Quick "SOHAC" for rushing up runtime.

(Amineh Amini et.al.) [1] acclimated a "novel clustering method" called DMM Stream for constant trajectory data. Undeveloped bit streams use this kind of clustering-based thickness calculation.

(B.A.Maroon, P. A. D. de Maine) [3] suggested the measurement of "DBCURE" to group extremely large data items. He uses a thickness-based classification method alongside parallel processing. In addition, the creators innovated a system map reduction called DBCURE-MR. All of the methods that have been presented are effective and can aid in the discovery of groups that are definitive for various data sets.

(V.C.Storey and I.-Y. Song) [26], The era of big data has resulted in the development and applications of technologies and methods aimed at effectively using massive amounts of data to support decision-making and knowledge discovery activities. In this paper, the five Vs of big data, volume, velocity, variety, veracity, and value, are reviewed, as well as new technologies, including NoSQL databases that have emerged to accommodate the needs of big data initiatives.

(V. Brock and H. U. Khan)[40], The big data concept, started applying the concept to various fields and tried to assess the level of acceptance of it with renown models like technology acceptance model (TAM) and its variations. By synchronizing TAM with organizational learning capabilities (OLC) framework. These models are applied on the construct, intended usage of big data and also the mediation effect of the OLC constructs is assessed. The study uses structural equation modelling and multivariate regression using ordinary least squares estimation to test the proposed hypotheses using the latest statistical software R. It is proved from the analysis that compared to other models, model 4 (which is constructed by using the constructs of OLC and TAM frameworks) is able to explain 44% variation in the usage pattern of big data. In addition to this, the mediation test performed revealed that the interaction between OLC dimensions and TAM dimensions on intended usage of big data has no mediation effect.

(J. S. Ward and A. Barker) [31], The term big data has become ubiquitous. Owing to a shared origin between academia, industry and the media there is no single unified definition, and various stakeholders provide diverse and often contradictory definitions. The lack of a consistent definition introduces ambiguity and hampers discourse relating to big data. This short paper attempts to collate the various definitions which have gained some degree of traction and to furnish a clear and concise definition of an otherwise ambiguous term.

(G. Manogaran, C. Thota and M. V. Kumar) [47], The size of the data that is generated grows exponentially. As cloud computing provides an optimized, shared, and virtualized IT infrastructure, it is better to leverage the cloud services for storing and processing such Big Data. Securing the data is one of the major challenges in all the domains. Though security and privacy have been talked about for decades, there is still a growing need for high end methods for securing the rampant growth of data. The privacy of personal data, and to be more specific the health data, continues to be an important issue worldwide. Most of the health data in today's IT world is being computerized.

(F. R. Damayanti, K. A. Elmizan, Y. F. Alfredo, Z. N. Agam and A. Wibowo) [27], In big data era, data may be in the form of

structured, unstructured and semi-structured. They become an asset for companies that contain important insights and support in decision-making processes. Big data is one of technologies that used by many organizations for managing store data and data analysis. Therefore, storage management is one of important processes in big data management to ensure that data are stored correctly, safely and easily to be accessed.

(A Singh, K Chatterjee) [28], The cloud computing provides on demand services over the Internet with the help of a large amount of virtual storage. The main features of cloud computing are that the user does not have any setup of expensive computing infrastructure and the cost of its services is less. In the recent years, cloud computing integrates with the industry and many other areas, which has been encouraging the researcher to research on new related technologies. Due to the availability of its services & scalability for computing processes individual users

(N Subramanian, A Jeyaraj) [30], Cloud computing is an archetype that enables access to a shared pool of computing resources for cloud users in an on-demand or pay-per-use, fashion. Cloud computing offers several benefits to users and organizations, in terms of capital expenditure and savings in operational expenditure. Despite the existence of such benefits, there are some obstacles that place restrictions on the usage of cloud computing. Security is a major issue that is always considered.

(Z. A. Al-Sai, R. Abdullah and M. H. Husin)[33], Big Data, their key characteristics, and the challenges surrounding the concept and characteristics. The implementation of Big Data brings new critical challenges that need to be addressed before starting the Big Data journey. The main challenges that are commonly faced by organizations are the implementation of Big Data alongside the critical challenges that involve the challenges related to technology, organization, process, data management, and skills. The challenges are difficult and the growth in terms of Big Data is increasing exponentially which calls for further investigation.

(N. Z. B. Zainal, H. Hussain and M. N. M. Nazri)[34], the age of big data. All over the world recently, the organizations including the government is taking initiative to adapt and implement big

data. Nowadays, government do not want to only collect data, but they want to understand the meaning and the significance of the data, as well as to analyse the data and aid them in making strategic decisions.

In 2013, big data defined by Gartner [35] as “high volume, high variety, high-velocity assets of information that require innovative Cost-effective forms of processing for intensifying insight discovery, for process optimization and decision making”.

(Hugh J. Watson) [36], the potential value of big data analytics is great and is clearly established by a growing number of studies.

The keys to success with big data analytics include a clear business need, strong committed sponsorship, alignment between the business and IT strategies, a fact-based decision-making culture, a strong data infrastructure, the right analytical tools, and people skilled in the use of analytics. Because of the paradigm shift in the kinds of data being analysed and how this data is used, big data can be considered to be a new, fourth generation of decision support data management.

(H. Braun) [32], Big Data is defined as high volume, high velocity and high variety information assets, a result of the explosive growth of data facilitated by the digitization of our society. Data has always had strategic value, but with Big Data and the new data handling solutions even more value creation opportunities have emerged. Studies have shown that adopting Big Data initiatives in organizations enhance data management and analytical capabilities that ultimately improve competitiveness, productivity as well as financial and operational results. There are differences between organizations in terms of Big Data capabilities, performance and to what effect Big Data can be utilized.

(S. M. Drus and N. H. Hassan) [38], Big data is one of technologies that used by many organizations for managing store data and data analysis. Therefore, storage management is one of important processes in big data management to ensure that data are stored correctly, safely and easily to be accessed.

(D. T. Widyaningrum) [39], Learning Organizations should consider big data as a strategic business asset. The benefits of big data analysis include increased cost efficiency, revenue growth, and competitive value in the market. Shorter product

cycles, global competition, and increased workplace differences are the factors that have triggered companies to adapt faster to challenging conditions. This article highlights the way in which big data affects an organization's success through the use of adequate analytics. Companies such as Amazon and Nike have proven the importance of using big data to support predictive and decision-making behaviours. It also affects the organization aspects such as strategy, people, structure, rewards, and process. It is not intended to deal with every element of the organization, but to explore the relevant relationships between big data and the learning process.

(R. Kumar and R Goyal)[44], The narrative review presented in this survey, however, provides an integrationist end-to-end mapping of cloud security requirements, identified threats, known vulnerabilities, and recommended countermeasures, which seems to be not presented before at one place. Additionally, this study contributes towards identifying a unified taxonomy for security requirements, threats, vulnerabilities and countermeasures to carry out the proposed end-to-end mapping. Further, it highlights security challenges in other related areas like trust based security models, cloud-enabled applications of Big Data, Internet of Things (IoT), Software Defined Network (SDN) and Network Function Virtualization (NFV).

(R. Velumadhava Rao and K. Selvamanib)[48], Cloud Computing trend is rapidly increasing that has an technology connection with Grid Computing, Utility Computing, Distributed Computing. Cloud service providers such as Amazon IBM, Google's Application, Microsoft Azure etc., provide the users in developing applications in cloud environment and to access them from anywhere. Cloud data are stored and accessed in a remote server with the help of services provided by cloud service providers. Providing security is a major concern as the data is transmitted to the remote server over a channel (internet).

(A.K. Chitturi and P Swarnalatha) [50], Cloud computing provides many aids to the users and organizations in accordance with expenditure and savings. Other than these benefits, cloud computing has some hurdles that result in restriction of its usage. Cloud security is the bigger hurdle which is regularly taken into consideration.

(P. R. Kumar, P. H. Raj and P Jelciana)[51], Cloud computing is one of the fastest emerging technologies in computing. There are many advantages as well few security issues in cloud computing. The data security issues in cloud computing in a multi-tenant environment and proposes methods to overcome the security issues. This paper also describes Cloud computing models such as the deployment models and the service delivery models.

(I. A. T. and Hashem)[42], Massive growth in the scale of data or big data generated through cloud computing has been observed. Addressing big data is a challenging and time-demanding task that requires a large computational infrastructure to ensure successful data processing and analysis. The rise of big data in cloud computing is reviewed in this study. The definition, characteristics, and classification of big data along with some discussions on cloud computing are introduced. The relationship between big data and cloud computing, big data storage systems, and Hadoop technology are also discussed. Furthermore, research challenges are investigated, with focus on scalability, availability, data integrity, data transformation, data quality, data heterogeneity, privacy, legal and regulatory issues, and governance.

CONCLUSIONS

We have reached to the time of data heavy downpour where data is main thing and it is available all over. Big data and its examination convince the perspective and business of person. Big data is the upcoming benchmark in the periods of data innovation. While big data innovates a new chance to fine pick up the new bits of knowledge of business, it additionally brings a few challenges which need to consider for accomplishing another constant advantages. This thesis characterizes the idea and attributes of big data and its difficulties with center around highlights of big data. This paper characterizes the big data challenges including related advancements. There is no uncertainty that big data innovation is still being developed, since all available big data procedures are constrained to take care of the big data issues totally. From equipment to programming, we despite everything need a progressively

advanced capacity and I/O methods to really tackle the big data issues.

REFERENCES

- [1] Amineh Amini, Hadi Saboohi, Teh Ying Wah, and Tutut Herawan, "Dmm-stream: A density mini-micro clustering algorithm for evolving data streams", Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013), pp. 675- 682, Springer, 2014.
- [2] Arthur R. Miller, Charles R. Ashman, The Assault on Privacy, 20 DePaul L. Rev. 1062 (1971).
- [3] B. A. Maroon, P. A. D. de Maine, Automatic data Compression, Communications of the ACM, Volume 10 Issue 11, November 1967 Pages 711-715.
- [4] Divyakant Agrawal, Big data and cloud computing: current state and future opportunities, EDBT 2011, March 22–24, 2011, Uppsala, Sweden. Copyright 2011 ACM 978-1-4503-0528-0/11/0003.
- [5] Dr. Saravana Kumar N M, Eswari T, Sampath P & Lavanya S, "Predictive Methodology for Diabetic Data Analysis in Big data, 2nd International Symposium on Big data and Cloud Computing (ISBCC'15), pp. 203-208, available at Science direct, 2015.
- [6] Ganesh Krishnasamy , Anand J Kulkarni, and Raveendran Paramesran, "A hybrid approach for data clustering based on modified cohort intelligence and k-means", Expert Systems with Applications 41, pp. 6009–6016, Elsevier, 2014.
- [7] Hong Yu, Zhanguo Liu, and Guoyin Wang, "An automatic method to determine the number of clusters using decision-theoretic rough set", International Journal of Approximate Reasoning, vol. 55, issue 1, pp. 101-115, January 2014.
- [8] Hongbo Zou, Yongen Yu, Wei Tang, Hsuan-Wei Michelle Chen, FlexAnalytics: A Flexible Data Analytics Framework for Big Data Applications with I/O performance Improvement, Special issue on Scalable Computing for Big Data Elsevier, Big Data Research, Volume 1, August 2014, Pages 4-13.

- [9] Istvan Dienes, A Meta Study of 26 “How Much Information” Studies, Sine Qua Nons and Solutions, Hungarian Central Statistical Office (HCSO) 1979–1997, *International Journal of Communication* 6 (2012), 874–906.
- [10] Krisztian Buza, Gabor I Nagy, and Alexandros Nanopoulos, “Storage optimizing clustering algorithms for high-dimensional tick data”, *Expert System with Applications*, vol. 41, issue 9, pp. 4148-4157, July 2014.
- [11] Marco Viceconti, Peter Hunter, and Rod Hose, Big data, big knowledge: “Big data”for personalised healthcare”, *IEEE Journal of Biomedical and Health Informatics*, vol. 19, issue 4, DOI:10.1109/JBHI.2015.240688, pp. 1209-1215, February 2015.
- [12] Mathur R, “Integrating Big data in Cloud Environment: A Review”, *International Journal of Innovations in Engineering and Technology (IJJET)*, volume 7, issue 1, pp. 513-517, June 2016.
- [13] Min Chen, Shiwen Mao, Yunhao Liu, *Big Data: A Survey*. Springer Science Business Media New York 2014.
- [14] Omer Tene & Jules Polonetsky, Privacy in the age of Big Data: A time for Big Decisions, Symposium Issue, 64 *STAN. L. REV. ONLINE* 63 February 2, 2012.
- [15] Ping Jiang, et.al, “An Intelligent Information Forwarder for Healthcare Big data Systems with Distributed Wearable Sensors”, *IEEE Systems journal*, January 2014.
- [16] Qing He , Xin Jin, Changying Du, Fuzhen Zhuang, and Zhongzhi Shi , “Clustering in extreme learning machine feature space”, *Neuro Computing*, 128: pp. 88-95, 2014.
- [17] R Madhuri, M Ramakrishna Murty, JVR Murthy, PVGD Prasad Reddy, and Suresh C Satapathy, “Cluster analysis on different data sets using k-modes and k-prototype algorithms”, In *ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-vol II*, pages, pp. 137-144. Springer, 2014.
- [18] Raghunath Nambiar, Adhiraaj Sethi, Ruchie Bhardwaj, Rajesh Vargheese, “A Look at Challenges and Opportunities of Big data Analytics in Healthcare”, *IEEE International Conference on Big data* , IEEE 2013.
- [19] Rohit Chandrashekar, et. al, “Integration of Big data in Cloud computing environments for enhanced data processing capabilities”, *International Journal of Engineering Research and General Science* vol. 3, issue 3, part-2 , pp.240-245, May-June 2015.
- [20] Ruchie Bhardwaj, Adhiraaj Sethi, Raghunath Nambiar, Big data in Genomics: An Overview”, *IEEE International Conference on Big data*, 10.1109/BigData.2014.7004392, 08 January 2015.
- [21] Sam Madden, From Databases to Big Data, Article, Massachusetts Institute of Technology, *IEEE Internet Computing* 2012.
- [22] Samuel Fosso Wamba , Shahriar Akter, Andrew Edwards d, Geoffrey Chopin, Denis Gnanzou, How ‘big data’ can make big impact : Findings from a systematic review and a longitudinal case study, *International Journal of Production Economics* 165 (2015)234–246.
- [23] Srivathsan M, Yogesh Arjun K, “Health Monitoring System by Prognostic Computing using BigData Analytics”, 2nd ISBCC 2015, *Procedia Computer Science* 50, pp. 602 – 609, available at Science direct, 2015.
- [24] Tetsu Akiyama, the continued growth of text information: from an analysis of information flow censuses taken during the past twenty years, *Keio Communication Review* No. 25, 2003
- [25] V Srinivas Jonnalagadda, P Srikanth, Krishnamachari Thumati, Sri Hari Nallamala, “A Review Study of Apache Spark in “Big data Processing”, *International Journal of Computer Science Trends and Technology (IJCST)* – vol. 4, issue 3, May - Jun 2016.
- [26] V. C. Storey and I.-Y. Song, "Big data technologies and Management: What conceptual modeling can do", *Data & Knowledge Engineering*, vol. 108, pp. 50-67, March 2017.
- [27] F. R. Damayanti, K. A. Elmizan, Y. F. Alfredo, Z. N. Agam and A. Wibowo, "Big Data Security Approach in Cloud: Review", *2018 International Conference on Information Management and Technology (ICIMTech)*, pp. 428-431, 2018.
- [28] Singh and K. Chatterjee, "Cloud security issues and challenges A survey", *Journal of Network and Computer Applications*, vol. 79, pp. 88-115, 2017.

- [29] S. Garg and M. V. A.K.R. Jyotsna, "Reviewing Security Concerns in Cloud Environment", *IJCSMC*, vol. 6, pp. 200-206, June 2017.
- [30] N. Subramanian and A. Jeyaraj, "Recent security challenges in cloud computing", *Computers & Electrical Engineering*, vol. 71, pp. 28-42, 2018.
- [31] J. S. Ward and A. Barker, *Undefined By Data: A Survey of Big Data Definitions*, 2013.
- [32] H. Braun, *Evaluation of Big Data Maturity Models - a Bench-Marking Study to Support Big Data Maturity Assessment in Organizations*, 2015.
- [33] Z. A. Al-Sai, R. Abdullah and M. H. Husin, "Big Data Impacts and Challenges: A Review", *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pp. 150-155, 2019.
- [34] N. Z. B. Zainal, H. Hussain and M. N. M. Nazri, "Big data initiatives by governments - Issues and challenges: A review", *Proc. - 6th Int. Conf. Inf. Commun. Technol. Muslim World ICT4M 2016*, pp. 304-309, 2017.
- [35] Gartner IT Glossary, *Answering Big Data's 10 Biggest Vision and Strategy Questions*, [online] Available: <https://www.gartner.com/doc/2822220?refval=&pcp=mpe#a-1319868613>.
- [36] H. J. Watson, "Tutorial: Big Data Analytics: Concepts Technologies and Application", *Communications of the Association for Information Systems*, vol. 34, 2014.
- [37] H. Macke, *Measuring Your Big Data Maturity*, 2017.
- [38] S. M. Drus and N. H. Hassan, "Big Data Maturity Model- a Preliminary Evaluation", *Proceedings of the 6th International Conference on Computing and Informatics*, pp. 613-620, 2017.
- [39] D. T. Widyaningrum, "Using Big Data in Learning Organizations", *Proc. 3Rd Int. Semin. Conf. Learn. Organ.*, vol. 45, pp. 287-291, 2016.
- [40] V. Brock and H. U. Khan, "Big data analytics: does organizational factor matters impact technology acceptance?", *J. Big Data*, vol. 4, no. 1, pp. 21, 2017.
- [41] G Firican, "The 10 Vs of big data", *Upside where data means business*, 2017.
- [42] I. A. T. and Hashem, "The rise of "big data" on cloud computing: Review and open research issues", *Information Systems*, vol. 47, pp. 98-115, 2015.
- [43] P.M. Mell and T. Grance, *The NIST Definition of Cloud Computing*, National Institute of Standards and Technology, September 2011.
- [44] R. Kumar and R Goyal, "On cloud security requirements threats vulnerabilities and countermeasures: A survey", *Computer Science Review*, vol. 33, pp. 1-48, 2019.
- [45] Yun Li et al., "Big Data and Cloud Computing", *Manual of Digital Earth*, pp. 325-355, November 2019.
- [46] *Hadoop and its relations with cloud*, June 2018.
- [47] G. Manogaran, C. Thota and M. V. Kumar, "meta cloud data storage Architecture for big data security in cloud computing", *Procedia Computer Science*, vol. 87, pp. 128-133, 2016.
- [48] M. Thilagavathi, D. Lopez and B. S Murugan, "Middleware for Preserving Privacy in Big Data", *Cloud Infrastructures for Big Data Analytics*, pp. 419-443, 2014.
- [49] R. Velumadhava Rao and K. Selvamanib, "Data Security Challenges and Its Solutions in Cloud Computing", *International Conference on Intelligent Computing Communication & Convergence*, pp. 204-209, 2015.
- [50] A.K. Chitturi and P Swarnalatha, "Exploration of Various Cloud Security Challenges and Threats", *Soft Computing for Problem Solving. Advances in Intelligent Systems and Computing*, vol. 1057, pp. 891-899, 2020.
- [51] P. R. Kumar, P. H. Raj and P Jelciana, "Exploring Data Security Issues and Solutions in Cloud Computing", *Procedia Computer Science*, vol. 125, pp. 691-697, 2018.
- [52] *Top Threats to Cloud Computing Tech. Rep. V1.0 Cloud Security Alliance 2010*, [online] Available: <https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf>.