

# Building a Big data provenance with its applications for Smart cities

V.M.Prabhakaran<sup>1</sup>, P.Devatharini<sup>2</sup>, L.Dineshkumar<sup>3</sup>, K.Prasanth<sup>4</sup>

<sup>1</sup>Assistant Professor, Dept of Computer Science and Engineering, KIT-Kalaignarunaidhi Institute of Technology, Tamilnadu, India

<sup>2,3,4</sup> Dept of Computer Science and Engineering, KIT-Kalaignarunaidhi Institute of Technology, Tamilnadu, India

\*\*\*

**ABSTRACT** - The sustainability of enormous and eternally data pools using different formats that cannot be processed with conventional software tools is the next big confront for web designers, Internet marketers and software engineers that requires new technologies and practices. One of the approaches to deal with Big Data is to use Semantic Web technologies, especially machine-interpretable metadata and Linked Data. Thus the advent of Big Data era has changed the outlook of numerous fields in science and engineering. We focused on Big data applications such as Healthcare, Education, Government, Media and Entertainment. This paper surveys real-time big data analytics applications and their technical challenges.

**Keywords:** Semantic Web technologies, Machine-interpretable metadata, Big data applications

## 1. INTRODUCTION

A Big Data is an area that holds a large of data or huge amount of data. Today we live in a world where Big data is driven everywhere. For instance, every 60 seconds there exists 98000+ tweets, 695000 status updates in Face book and soon that extend to about 1,820 TB of data. Not only social media, it is also being used in various fields like Education, Banking, Healthcare, Industrial. It is a term with both structured and unstructured data that includes obtaining data, analysis, transfer, editing, privacy and so on. Big Data analytics is classified accordingly Cost reduction, Decision making and New product/services.

**COST REDUCTION:** It involves the cutting of cost from expected level. An example for this is, Intel use Big Data analytics for making chips. In a brief there was a requirement for 19000 tests to be conducted on the chips , but with Big Data analytics process ,the tests were focused

only on specified chips to cut down the test time , which saved Intel \$3 Million in manufacturing.

**DECISION MAKING:** It helps the companies to make improved and clear decisions. The obtaining data undergo various steps like formatting, transfer, storage and analysis for focused improvement.

**NEW PRODUCT/ SERVICES:** It uses predictive analysis to introduce new products in the market. Firms can create product that connect with the customer, which provide increased customer value and minimize the risks associated with the launch of a new product. For instance, P&G uses this term and focuses to launch a new product when there is a need for it.

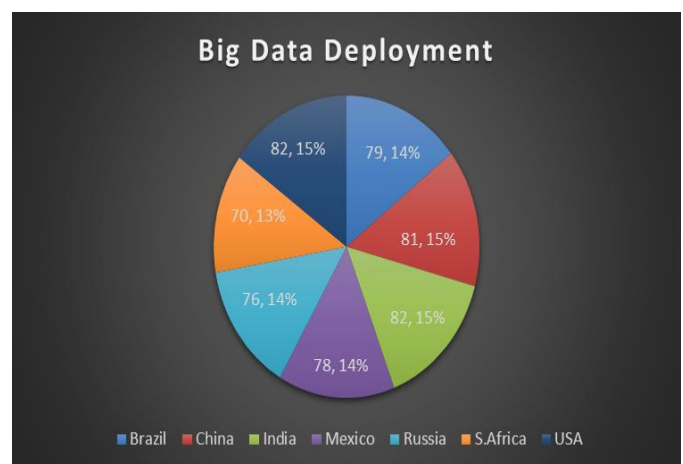


Fig-1: Big Data Deployment

## 2. LITERATURE SURVEY

Table -1: Literature survey of Big data analysis of past, present and future

S. No.	Paper title	Authors	Problem statement
1	Big Data Driven Optimization for Mobile Networks towards 5G	Kan Zheng, Zhe Yang, Kuan Zhang, PeriklisChatzimisios, Kan Yang, and Wei Xiang	Big data convenience in mobile network for treasuring immeasurable data.
2	Recent achievements and new challenges	Gema Bello-Orgaz , Jason J. Jung ,David Camacho	Big data is a far reaching excessive number of scrutinizes depleted in social media for information compound.

3	Understanding how big data leads to social networking vulnerability	Romany F. Mansour	To notorious entangled social junction while the big data running.
4	Big data driven smart energy management: From big data to big insights	Kaile Zhou, Chao Fu , Shanlin Yang	Big data managed distinct technologies for smart energy authority.
5	Big Data Meet Green Challenges: Big Data Toward Green Applications	Jinsong Wu, Senior Member, IEEE, Song Guo, Senior Member, IEEE, Jie Li, Senior Member, IEEE, and Deze Zeng, Member, IEEE	Big data is immensely profitable at foliage thread for future propagation.
6	Big data driven smart energy management: From big data to big insights	Kaile Zhou, Chao Fu , Shanlin Yang	Scrutiny of human emergency the smart clothing is advanced by mingling human with cloud
7	Big data driven smart energy management: From big data to big insights	Kaile Zhou, Chao Fu , Shanlin Yang	Big data is an enormous amount of data to renders smart energy supervision.
8	Big-data for building energy performance: Lessons from assembling a very large national database of building energy use	Paul A. Mathew , Laurel N. Dunn, Michael D. Sohn, Andrea Mercado, Claudine Custudio, Travis Walter	Big data for building energy has deplete to posses energy flowing building for panaroma size and application.
9	Big Data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook	Radu F. Babiceanua, RemziSeker	To syndicate a world with cyber circulating world for future panorama .
10	Big data analytics based fault prediction for shop floor scheduling	Wei Ji, LihuiWanga,	Big data appliances journal erection system for deferment provinces.
11	Big Data Research in Information Systems: Toward an Inclusive Research Agenda	Ahmed abbasi, SuprateekSarker, Roger H. L. Chiang	Big data probing in agenta for perception between big data future.x
12	MERRA Analytic Services: Meeting the Big Data challenges of climate science through cloud-enabled Climate Analytics-as-a-Service	John L. Schnase , Daniel Q. Duffy , Glenn S. Tamkin , Denis Nadeau, John H. Thompson , Cristina M. Grieg , Mark A. McInerney , William P. Webster	Big data disputes in climate through cloud computing to provide new category in the data ceraminal stack
13	MetaCloudDataStorage Architecture for Big Data Security in Cloud Computing	GunasekaranManogarana, ChanduThotab,M. Vijay Kumarc	Big data safeguard in cloud to advance metadata .
14	Real-time business activity monitoring and analysis of process performance on big-data domains	Alejandro Vera-Baquero ,Ricardo Colomo-Palacios, Owen Molloy	Big data depleted in business enterprise for concurrent destination.
15	Intelligent cryptography approach for secure distributed big data storage in cloud computing	Yibin Li , KekeGai, Qiuc, MeikangQiu, Hui Zhao	To safeguard data privacy has cynical in cloud so privilege numerous path created for web employment.

### 3. APPLICATIONS OF BIG DATA

The various applications of Big data are,

- A. Health Care
- B. Erection Of Big Data

- C. Big Data Enlargement
- D. Government Proving Big Data
- E. Media & Entertainment
- F. Education

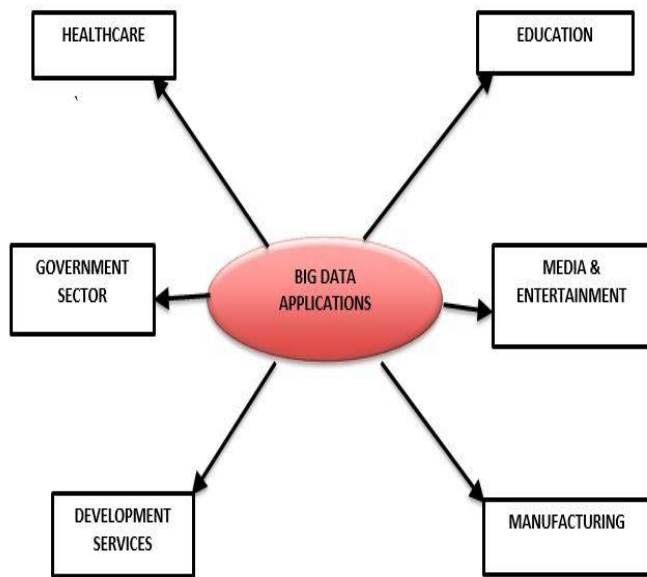


Fig-2: Applications of Big Data

**A. HEALTH CARE**

In health care industry the data can be generated and stored as hard copy it has the capability to support a medical functions. Big data analytics has become the crucial problem in healthcare informatics such as data querying, clinical decision support etc., Big data is originally generated in large electronic data sets, to manage the data with conventional software or hardware. Over the past decagon electronic health care records(EHR) has been adopted in clinics, patient disease pattern can be studied in such data it will helpful to improve patient care efficiently. Also uses "sniffer" algorithms to identify the patient risks. "Health data allows the doctor to shape the patient profile effectively presumes and treats disease".

In traditional system of data storing makes simple with the help of medical transcription workers, big data replaces it and storing digitally arrive of big data technology healthcare industry is going liberal of some application like health care intelligence, monitoring patient vitals etc., it reduces the cost burden appreciate the quality of human life by halt the death. Researchers of 80% the health care data becomes indeterminate and great challenges for the industry. Due to this technology the industry spread their wings and to support the healthcare management function such as disease surveillance, population health management.

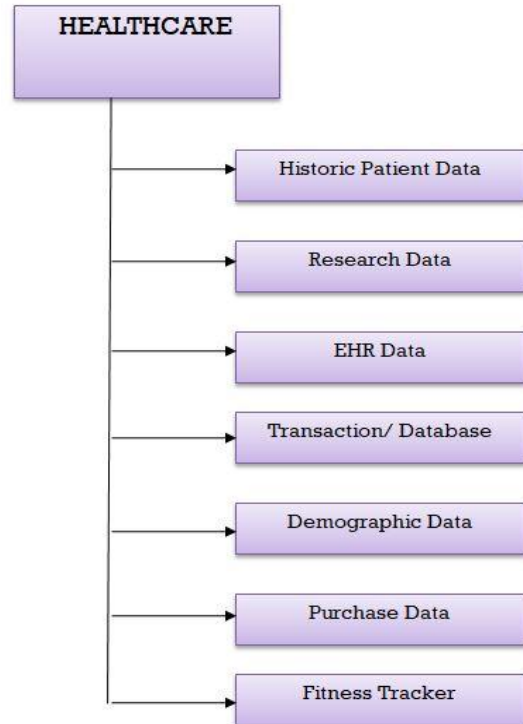


Fig-3: Representation of Healthcare Process

"Healthcare cost is increasing day by day so we demands of some motivated thinking in this field". In this industry the problem isn't the lack of data it's used to support some decision making, planning etc., For multi-patient, it's tricky to generate and lead to the emergence of big data technology.

When coming to the Smart Healthcare Protection Idea, if an accident happens in the road or in any public place the alarm automatically intimates the emergency ambulance service and it quickly reach the accident spot within the area.

**B. ERECTION OF BIG DATA**

Manufacturing organizations are glancing the ways to improve their prediction while reducing costs defect tracking to recuperate overall efficiency of improvements. Customer experience is favored positively in the operational improvements, maintaining new awareness information is discerning. An assembling companies become larger the data must be handled more composite. In social media the external data can be supplied the network channels will implement a huge penetration. With the help of big data solutions the operations can lead to improved and timely decision making.

Today the Big Data Root Can Accommodate

- (i). Operational system from conventional scheme
- (ii). Chain data from sensor in manufacturing
- (iii). Data from marketing

(iv). Sales offensive

(v). Announcing retort

Mass production companies have long lists of production and conveyance measures. To salvage the enforceable information from flourishing of data. In an organization the list of top patches in data technologies: Products of reformed telegraphing and production, real time decisions. A new way of customer support is an customer experience. Developed information can qualify to serve their customer in discursive way. In past 20 years to enhance the legacy quality and revenue. The big data era has originated while the advanced analytics is grounded in years of mathematical research and scientific application. The biopharmaceutical production that harvest can differ from 50 to 100%.

A company saves annually by 50 % resulting in between \$5 and \$10 million of the aiming process. At some time the Intel has exploiting big data for the processor. The chipmaker tests the chip at staging line. The processor of individual Intel core the company should save \$3million for completion cost. For manufacturing a chip the company can save an appended at \$30 million due to the widening of big data. Due to the completion the big data can spell out the any problem.

**C. BIG DATA ENLARGEMENT (Development Services)**

"Big data is a documentation of source of big data pertinent to the policies and planning for enhancement program". It's totally differing from the "Traditional development data". Big data achievements are those that can be estimate to lecture insight into human well-being and augmentation.

In general, provenance of big data provides important benefits in key development patches. These patches consider resource management, economic productivity, health care, natural disaster, job marketed. Big data will be the fuel dissemination to next industrial reformation radically renovation of economic structure it's reaching into every aspects of social life. In 1946 the first computer weight is thousands of kilograms but it does not higher than 500 calculations per second these days, the IBM Watson supercomputer can manner 500 gigabytes per second that is the correlative of reading one million books per second. Big data should categorize into two concepts: Information and Communication Technology (ICT).

It precedes the exponential potential for international development work across collective fields. There is no end of these technologies that are qualifying the use of big data. The growing of mobile phone technologies has affected the developing countries in the past tenner. It has been consumed to transfer money, test results in daily life etc. Uses of big data in mobile analytics to collect the call volume, calling pattern, and location call detail records etc. Visual analytics are the most impressive side of big data

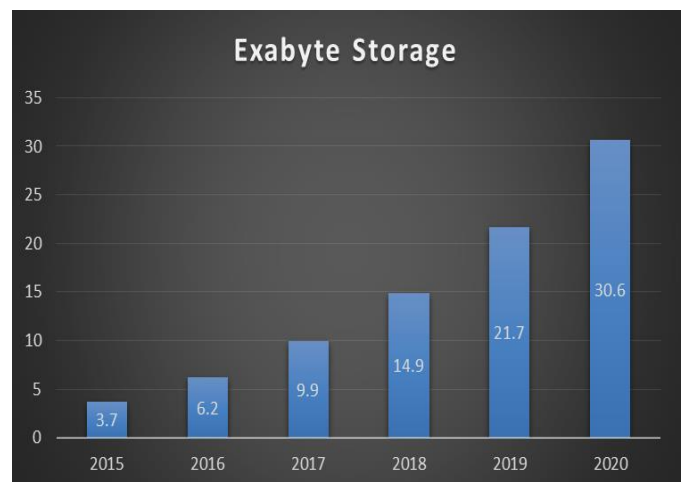
exploration to support the science of analytical reasoning through interactive visual interfaces, in an limited size the qualified information can be viewed. Although the visual analytics does not prolong the information about the human interface, quickly recover the data to the user converse to automatic data mining and machine gathering tools it require soundness and knowledge.

Mobile technology dataset is useful for studied about the social field to know more about the knowledge of social network analytics these technology can be utilized to appreciate multiple tools.

**D. BIG DATA IN GOVERNMENT SECTOR**

Government employing big data to strengthen their ability to serve citizens and also involving the economy, health care, job creation and terrorism. Big data helps the government to abundance and influence in security, substantial for other needs of people. Big data is an unorganized analysis it is too large and raw techniques. Big data is a large amount of digital data and possessed from all foundations. "Around 90% of the world has been accomplished due to this past two years with 2.5 quintillion bytes of data added every day, 90% are unregulated".

In any organization the big data in cloud, tenders more openings to discovery, value creation and business intelligence for decision support. It also has new protests for complexity, security, and risks to privacy as well as occasion for new technology and human skills. The data can be redefined by management system from extracting, transforming, loading into new technologies. Although big data development the business sector is a leading position, and the public sector helps to take the decision shaping in real-time from fast growing in motion data from multiple parts including the web, video, email and social communications. To serve their citizens by the help of articles and business reports from the natural disasters, healthcare cost and terrorism. To transforming of data the government must proceeds new technologies.



**Fig-4:** Storage analysis of Big Data



It makes difficult to monitor while the criminal attempts heavily in sophisticated technology and innovations. The new York city in the early 1990s, the crime rates are extremely increased and the different police forces was selected to explore how big data can be influenced to anticipate and limited the crimes. The number carnages between 1985 and 1990 has been increased from 1,392 to 2,262 and only five years the 60% has increased. Using NYPD to hinder the high crime rates and accumulating the citizen security.

During big data analytics the government has assessment of dealing in the country. While the Indian government is also use the big data for income tax estimations. CCTV cameras and satellite are growing more data to the social environment. The government spreading big data to secure and maintain the individual sanctuary.

## E. MEDIA & ENTERTAINMENT

The media and entertainment industries have been always at the forefront of adopting new technologies. The problems that are driving media to look at big data capabilities are the need to reduce the costs of operating in an increasingly competitive landscape and, at the same time, there is a great need to analyze and store the data generated in diverse platform. In the field of media, the people in that field are connected with their customers and competitors than ever before. Data can be generated, shared to anyone all over the world with an Internet-enabled device. Global revenues from such devices, including Smartphone's, tablets, desktop PCs, TVs, games consoles, e-readers, wearable gadgets, etc... This means that the ability of big data technology to ingest, store, and process many different data sources, and in real-time, is a valuable asset to the companies who are prepared to invest in it. Analogue products such as newspapers have been created through digital technologies. Unlike some other sectors, the vast majority of actionable data in the media sector is already in digital form.

The media sector always generates huge amount of data, whether from research, sales, customer databases, log files, and so on. The big data provides technical solutions, and strategies offer the ability to manage and disseminate data at speeds and scales that have never been seen before. Big data is also useful in analyzing data not only at the customer and products levels, but also at network and infrastructure Big data plays an important role in processing and storing and organizing data that are generated by people in media field.

Thus, the advantages of big data are explained in five scenarios. They are:

**(i). Data Journalism:** Large volumes of data have been developing by media and also for the media. Single or multiple datasets require analysis to derive insight, find interesting stories, and generate material.

**(ii). Dynamic Semantic Publishing:** Scalable processing of content for efficient targeting. Using semantic technologies to both produce and target content more efficiently.

**(iii). Social Media Analysis:** Processing of large user-generated content datasets. Real-time analysis of millions of tweets, images, status, videos, messages, updates to identify trends and content that can be packaged in value-added services.

**(iv). Product Development:** Huge generation of data for the analytics of new services for the development of products.

**(v). Audience Insight:** Using data from multiple sources to build up a comprehensive 360 view of a Customer. Entertainment and Social media companies should analyze data that should not be in customer and product levels alone, it should cover the network and infrastructure levels.

## F. BIG DATA IN THE FIELD OF EDUCATION

In the last 20 years, the computing technology has been growing fast replacing all the traditional methods. One of the most important missions of computing is to share the data. Business and government agencies are finding the strategic uses of large databases. New software tools and techniques are assumed to analyze the data for beneficial inferences.

Even more, the web establishes new needs for Information development. There is an enormous increasing in the users level as well as information level leads to web research. People used to surf the web always as it become a regular habit of them starting with high quality human maintained indexes such as Yahoo! or with search engines, like Google. Big Data can provide advanced parallel techniques, such as, data Analysts (both human and machine), massive swaths of data to forecast situations like activities and procedures. Hence, Big Data has developed the estimates, save money, boost the productivity etc...It also improve decision-making in fields such as traffic control, weather forecasting, disaster prevention, finance, fraud control, business transaction, national security, health care and Education. Finally as we can see from the above, all have as a result many basic changes with the data. Creating a new era in which all processes and interactions including scientific research are affected. In this case Big Data can provide more opportunities for new learning experience for children and young adults. Hence students can share lot of information among themselves and also with their with educational institutions, in this way they can expand their knowledge and skills. Furthermore, Educational institutes and Universities are able to help and prepare their future.

Big Data plays really an important role in improving the education and can afford to shape a modern and dynamic education system, which every individual student can have the maximum benefit from that. Furthermore teachers have valuable tools, were they do not have before, which can

make their decisions more specific and are able to choose a big variety of new learning methods. Hence the Big Data are actually involved to change the way of industries including the education. Due to advanced handling techniques of big data the traditional difficulties will be no longer exists, keeping the good methods. The education system will be enriched with new learning ways, making more efficient and targeted.

#### 4. CONCLUSION

In the past, Big data have been used with predictive analytics, high performance computing systems, machine learning and other strategies and will continue to be used heavily in the future of all the fields. Huge amount of data can be stored, processed, accessed, and retrieved easily with the help of big data. By using the big data related systems, Engineers and Scientists have been able to discover new things and they have also been able to more accurately predict daily weather as well as natural disasters.

#### ACKNOWLEDGEMENT

V.M.Prabhakaran, P.Devatharini, L.Dineshkumar and K.Prasanth wishes to thank the Management, the Director, the Principal, of their institute, KIT-Kalaignarkaranidhi Institute of Technology, for providing all the necessary facilities and never ending support for the work. We wish to thank their Department Head, Prof.Dr.R.Sukumar for the freedom to pursue research and excellent research ambience provided by him. Our special thanks go to, Prof.Dr.R.Nedunchezian, Prof.Dr.S.Balamurugan and Prof.Dr.P.Raviraj Pandian for sowing the seeds of thinking big in research, their expert guidance and continuous motivation.

#### REFERENCES

- [1]. Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., & Ratti, C. (2011). E. Baralis, L. Cagliero, T. Cerquitelli, P. Garza, and M. Marchetti, "Cas-mine: providing personalized services in context-aware applications by means of generalized rules", Knowledge and information systems, vol. 28, no. 2, pp. 283-310, 2011. Real-time urban monitor using cell phones: A case study in Rome. Intelligent Transportation Systems 12(1), 141-151.
- [2]. S. Pandey, W. Voorsluys, S. Niu, A. Khandoker, and R. Buyya, "An autonomic cloud environment for hosting ecg data analysis services", Future Generation Computer Systems, vol. 28, no. 1, pp. 147-154, 2012.
- [3]. Laursen, K., & Salter, A. J. (2014). A. Ibaida, D. Al-Shammery, and I. Khalil, "Cloud enabled fractal based ecg compression in wireless body sensor networks", Future Generation Computer Systems, vol. 35, pp. 91-101, 2014.

- [4]. Dameri, R.P.: Comparing smart and digital city: initiatives and strategies in Amsterdam and Genoa. Are they digital and/or smart? In: Dameri, R.P., Rosenthal-Sabroux, C. (eds.) SmartCity. How to Create Public and Economic Value with High Technology in Urban Space, pp. 45-88. Springer, Heidelberg (2014).
- [5]. Sankaranarayanan S. Balamurugan, Dr. P. Visalakshi, V. M. Prabhakaran, S. Charanya Strategies for Solving the NP-Hard Workflow Scheduling Problems in Cloud Computing Environments. Australian Journal of Basic and Applied Sciences (2014).
- [6]. V.M. Prabhakaran, Prof. S. Balamurugan, A. Brindha, S. Gayathri, Dr. Gokul Kruba Shanker, Duruvakkumar V.S. NGCC: Certain Investigations on Next Generation 2020 Cloud Computing-Issues, Challenges and Open Problems Australian Journal of Basic and Applied Sciences (2015)
- [7]. V.M. Prabhakaran and Dr. Gokul Kruba Shanker S. Balamurugan, R.P. Shermey Internet of Ambience: An IoT Based Context Aware Monitoring Strategy for Ambient Assisted Living. International Research Journal Of Engineering and Technology (2016)
- [8]. Bencardino, M., Greco, I.: Smart communities. Social innovation at the service of the smartcities. TeMA. J. Land Use Mob. Environ. (2014)
- [9]. Alexopoulos, C., Zuiderwijk, A., Charapabidis, Y., Loukis, E., & Janssen, M. (2014). P. Neirotti; A. De Marco; A.C. Cagliano; G. Mangano; F. Scorrano (2014). Current trends in Smart City initiatives Designing a second generation of open data platforms: integrating open data and social media. Electronic Government (pp. 230-241). Berlin Heidelberg: Springer.

#### BIOGRAPHIES



**V.M. Prabhakaran** obtained his B.E degree in Computer Science and Engineering from Hindusthan Institute of Technology, Coimbatore, Tamilnadu, India and completed his M.E degree in Computer Science and Engineering from Kalaignarkaranidhi Institute of Technology, Coimbatore, Tamilnadu, India. He has to his credit 40 papers in National/International Journals/Conferences. He is the recipient of gold medal and certificate of merit for best journal publications by his host institution for the year 2013-14. He served as a Secretary for CSE Association at Hindusthan Institute of Technology 2012-13 also served as a President for CSE Association at Kalaignarkaranidhi Institute of Technology 2014-15. He is an active member of ISTE, IAENG, IRDP. Completed consultancy project regarding Medical Healthcare. Acted as a coordinator for various evaluation process NBA, NAAC, NIRF and MDRA. His area of research includes Cloud computing, Big data analytics and Cyber Security.



**P. Devatharini** is pursuing her B.E. degree in Computer Science and Engineering at KIT-Kalaignarkarunanidhi Institute of Technology, Coimbatore, Tamil Nadu, India. She completed her project in Augmented Reality. She has certified for attending the Two days conference on "NATIONAL CYBER SAFETY AND SECURITY'17" organized by National Cyber Safety and Security at BITS Pilani-Hyderabad and attending "SECURITY SUMMIT' 17" organized by Cloud Security Alliance at Dr. G R Damodaran College of Science, Coimbatore. She has certified for attending one day workshop on "Ethical Hacking" at KIT-Kalaignarkarunanidhi Institute of Technology, Coimbatore. She has certified for her 24 hours involvement in "THE LARGEST LEARNATHON OF INDIA", conducted by ICT Academy and Trailhead. She Presented his project model related to Augmented Reality and Won 2<sup>nd</sup> Place in the Mini-Project Expo organized by Department of Computer Science and Engineering in his host institution. Her areas of research includes Internet of Things(IoT), Computer graphics, Augmented reality and Network security.



**L. Dineshkumar** currently pursuing his B.E degree in Computer Science and Engineering at KIT-Kalaignarkarunanidhi Institute of Technology. He is serving as a Treasurer for CSE Association. He is also the student co-ordinator for organizing various events. He Presented his project model related to Ethical Hacking and Won 3<sup>rd</sup> Place in the Mini-Project Expo organized by Department of Computer Science and Engineering in his host institution. His area of interests include web designing , Big data analytics and Interested in Hacking related Technologies.



**K. Prasanth** currently pursuing his B.E degree in Computer Science and Engineering at KIT-Kalaignarkarunanidhi Institute of Technology. He is serving as a Joint Secretary for CSE association. He is also one of the student co-ordinators in his college and organized various events. He is one of the member in NCDRC (National Cyber Defence Research Centre) in his Institution. He has participated in the National Cyber Safety and Security Standards Summit '17 at BITS Pilani-Hyderabad Campus. Also participated in various ICT Academy organized events like Youth Leadership Summit '17 and New India Student Championship -2018(Largest Learnathon Of India). His area of interests include web designing, Big data analytics and Augmented reality.