Big Data in IoT

Conference Paper · July 2019

DOI: 10.1109/ICCCNT45670.2019.8944495

2 authors:

Shivanjali Khare
University of New Haven

Michael Totaro
University of Louisiana at Lafayette

Big Data in IoT

1st Shivanjali Khare
University of Louisiana at Lafayette
Center for Advanced Computer Studies
Lafayette, USA
sxk7139@louisiana.edu

2nd Michael Totaro
University of Louisiana at Lafayette
Center for Advanced Computer Studies
Lafayette, USA
mwt3774@louisiana.edu

***Abstract*—The Internet of Things is generating an enormous amount of data. Analyzing and managing that data requires programming and statistical approaches. Big Data technology operates on this massive data and pushes new products, applications, future research and developments to improve decision making. In this paper, we explore Big data in IoT driven technologies and the issue of the four V's in Big Data. This paper also highlights the importance of pre-processing, meta-data, data storage formats, data management and how big data is closely associated with IoT technologies. Today, with the rapid growth of IoT, everything is connected. To stay ahead of demands, new technologies such as Cloud Computing and Edge Computing are transforming IoT organizations. This paper discusses in which layers edge computing operates in the IoT reference model to achieve low-latency and greater efficiency solutions. This paper also reviews the IoT reference model layers that are associated with cloud computing, the structure of cloud computing architecture, data acquisition and data cleaning. This paper also discusses on various cloud-based IoT platforms such as AWS, Google Cloud IoT, Microsoft Azure, and Cisco IoT Cloud. We examined the importance of Big Data visualization, gives insights on various visualization tools and techniques. Lastly, this paper also addresses various significant challenges of Big Data in IoT, security issues and future research directions.**

***Index Terms*—IoT, Big Data, IoT security, meta-data, pre-processing, Edge Computing, Cloud Computing, Data cleaning, data acquisition, data visualization.**

I. Introduction

Internet of Things (IoT) is generating massive quantities of data every second. Bernard Marr in [1] projects the increase in data creation from past years. The Internet daily generates a massive amount of data through various services such as web searches, social-media platforms such as Facebook, Instagram, and so on. IoT is accelerating these statistics by connecting physical devices (sensors) to the Internet, providing variety of services to its users, while collecting different kinds of data. IoT involves data management and data analysis techniques. Data analysis requires an exclusive approach. Many organizations accomplish the data generated from IoT devices and use these insights for smart decision-making. Kashmir Hill in [2] cites an example where a US-based store, Target, was able to detect the pregnancy of women with advertising and purchases they made through credit card and analysis of their routine purchases against historical data.

IoT has many applications such as in healthcare, manufacturing, industrial IoT, smart homes, smart cities, and so on. IoT devices require the right form of sensors to be deployed in the right areas to capture the data. The collected data can vary, depending upon the service provided by the IoT device. IoT sensors have few restrictions such as environment sensitivity, distance limitations, etc. IoT sensors gather information from the environment, forwards it to the central node where data analysis take place, and then forwards the information to another node. Consider a smart home, which consists of multiple IoT devices such as thermostats, smart lighting systems, smart door locks, smart gardening, personal assistants, and so on. Across the entire house, there are bundles of nodes passing formation to the main server which stores or communicates this information the cloud. The user should be aware of restrictions by the sensors, which affect the data analysis, in order to avoid inaccurate or bad data.

The use of IoT devices shows a continuous collection of data. Gathering this data leads to observations that are remarkable. Big Data deals with the data set, analyzes and extracts meaningful information from collected data. There exist various online sources that provide open access data collections [3] [4].

The objective of this paper is to highlight the association between Big Data in IoT and create a relationship that determines the processing and analysis of data collected by IoT devices. This paper discusses big data management techniques at various levels such as collection, processing, analysis, and so forth. This paper also provides a survey of the existing IoT related technologies such as cloud and edge computing. The paper also delivers many attributes that are not addressed in current survey papers, along with some new challenges and future research.

II. Related Work

A review of the IoT literature suggests that there is considerable eagerness in the field of IoT systems [5] [6] [7] [8]. These studies, however, centered their research direction entirely on the architecture, applications and investments. Yunhao et. al in [9] reviewed the state-of-the-art of big data. They introduced general background, examined several applications and related technologies. Archenaa et. al in [10] focus on the seriousness of performing big data analysis on the data collected by the healthcare and government. In contrast, our work focuses on the techniques and mechanisms of data collected by the IoT devices and establishes a correlation between them. Figure 1 illustrates topics covered in this paper.

Discussed Topics						
Big Data	Intelligent Data Processing	Data Storage formats and databases	Data management	Data at the edge	Data in the cloud	IoT and Big data Visualization
•Volume •Variety •Velocity •Veracity	•Pre-processing •Meta-data creation	•Structured databases •Unstructured data storage	•IoT device security	•IoT reference model •Data acquisition	•IoT reference model •Data cleaning •Why is data stored in cloud? •Cloud architecture •Cloud based IoT platforms	•Data Visualization Techniques •Data Visualization Tools

TABLE I
STRUCTURE OF THE PAPER

III. BIG DATA

Gathering such massive data integrates storing the data generated from multiple technology nodes. IoT networks operate depending upon the analysis of this data. The network generate different data types, noise and some redundant data. Since IoT sensors and devices degrade over time, Big Data organizations reduce the risk of errors and maintain accurate decision making. Wetzkar et al. in [11] show various examples in the area of Industrial IoT (IIoT) where they faced issues in identifying, analyzing failures and troubleshooting failures. There is the need of automatic data collection and automatic error corrections. Traditionally, big data involves four dimensions, also known as Four V's. They are:

1) Volume: amount of data
2) Variety: different types of structured and unstructured data
3) Velocity: processing speed of the data
4) Veracity: truthness of the data

Some research scholars list these issues of Big Data as 3Vs' by removing Veracity, or by adding more issues such as Value, Validity, and so on. Ishwarappa and Anuradha in [12] considered 5Vs', whereas Khan in [13] considered 10Vs' as Big Data issues.

A. Volume

IoT devices stores massive data such as employee records, stock information, invoices, purchase history, card details, along with location details, and so on. Such additional information is called a meta-data that helps to contextualize the knowledge. The majority of large organizations invest in cutting edge databases, data management firms, distributed systems, and cloud storage for storing digital information. The quantity of data generated and collected from IoT devices is essential as all the data needs to be measured, stored or transmitted to other nodes. This has become a challenge as the amount of data has become very large and traditional database technology is no longer favorable.

B. Variety

Big data involves the gathering of target data from a wide range of sources simultaneously. IoT data involves data from different kinds of sensors, non-numerical items such as mp3, mp4, radio signals, and so on. Handling this variety of data is a challenge. The meta-data should be stored in correct context with the collected data and should allow to associate future data collections automatically. Another issue when considering the current state-of-art of IoT and change in techniques is the ability for storage software to adapt to these changes. For example, change of video quality or format in sensors.

C. Velocity

The data produced by sensors or other inputs in IoT devices occur at an extremely high rate. This high velocity of data production and collection becomes challenging because the data should be handled promptly for new data to come in. Moreover, the velocity of data production is not always constant. The velocity changes over time; for example, sales of a company increases during a certain offer period. Gandomi and Haider in [14] discuss the importance of time here. In such situations, there is a need for appropriate planning, processing power and storage to avoid data loss and system outage. Although such a commitment of computing power may be expensive, it should be planned ahead of time to increase the revenue of an organization.

D. Veracity

IoT sensors do not have margins of error in measurement. Wireless sensors can face communication error, hardware failure due to shift in the environment, animals or any other factors. As such, it is essential that data is properly stored, accurate and complete. The "truthness" of data forms the basis of many business decisions. It is necessary to differentiate between reliable and unreliable data.

IV. INTELLIGENT DATA PROCESSING

One common solution to the problem encountered during data collection and use of big data in IoT is the intelligent use of software. Some general approaches of intelligent data processing are Pre-processing and Meta-data creation.

A. Pre-processing

The data collected by IoT sensors is often sent to different locations and processed there. The large amount of data produced needs to be sent quickly to the processing location. Data can be lost entirely or in part if there is latency. Baker et al. in [15] discuss instances of medical emergency situations where such delays in communication can lead to possible detrimental effects on patients. Often, in many situations the data regarding particular event is required for further processing. Pre-processing helps to reduce the volume of data. It moves the processing function closer to the sensors and reduces the amount of data to be sent. Smart sensors in IoT uses built-in resources to perform pre-processing before sending it further. Antonini et al. in [16] presents a design framework for smart audio sensors. These smart sensors locally perform the computations on raw audio streams before transmitting those features wirelessly to IoT gateway.

B. Meta-data creation

After processing, data is stored to be used again. Meta-data is used to put the stored data into context. When needed the stored data is queried for information. Given the variety and volume of data, it can take a considerable amount of resources to process that data again. Meta-data to speed up the process by adding additional data that describes or references the stored data. Park et al. in [17] proposed a conceptual meta-data model for sensor data abstraction in IoT environments. This model helps to create a structured format for the low-level context and helps in higher abstraction procedures. Dawes et al. in [18] describe a deployable system to bridge the gap between data management. They propose a tiered meta-data recording system using a non-semantic and a semantic wiki related to a single sensor. Stevens in [19] discuss the importance of meta-data in big data analytics.

V. Data storage formats and databases

The relational database is used in traditional technical environments to store the data. They are used extensively and dominate most of the commercial data storage. The characteristics of IoT data make the traditional relational-based data management impractical. The use of a relational database can make the overall querying slow and might result in delayed responses.

A. Structured databases

An IoT program can be made more flexible by involving few restrictions, but it often makes the system less efficient. It is necessary to consider trade-offs while developing an IoT system. Relationship among the data elements establishes the structure of a database, making it efficient for storage and querying. The structured database leads to a lack of flexibility with modern software methodologies.

IoT devices have achieved technical advances and are able to communicate with almost any "thing." It requires an expansion of a network to accommodate more devices and their software. This is known as horizontal scalability. With the relational database, it becomes difficult to break these multiple clusters of machines. Sarkar et al. in [20] proposed an architecture to tackle the issue of scalability.

B. Unstructured data storage

Modern data today has made relational data management less efficient. Unstructured (also referred to as document store) and Semi-structured databases are developed to meet the needs of different types of data collected by IoT devices. Kumar in [21] discuss various techniques for maintaining unstructured data in IoT. According to Alnsari et al. in [22], due to massive developments in information technology, there is a need for solutions that should enable unstructured data management and analysis .

A new range of databases such as MongoDB and NoSQL are becoming more significant in IoT developments. They are unstructured database platforms that are proven effective in many IoT applications. NoSQL is also a non-relational database that can efficiently store key-value pairs, wide columns or search engines data, and so on. It makes them ideal for big data use and in particular IoT device development. Serdar in [23] discusses NoSQL in detail and outlines the advantages such as flexibility and overcoming horizontal scalability in detail.

VI. Data Management

Collecting and utilizing data can be useful but it also carries many risks and responsibilities. There are legal and ethical issues involved in collecting data without consent. This results in data breaches, which damages individuals' privacy. Guan et. al in [24] discuss how hackers can access the IoT data by multiple sources and use it for illegal benefits.

A. IoT device security

Many IoT devices that are accessible via the network should have some sort of credentials by which to connect. Unfortunately, this is not the situation. Many IoT devices are shipped without authentication to connect with or have default credentials which are highly insecure. In many situations, those devices that come with complex authentication details do not include credential changing manual which makes them vulnerable to attack once the credentials become known. IoT devices have thus become ideal targets for hackers. With the growing number of IoT devices, there is an increased risk of attackers present in a botnet. IoT devices can be used for multiple functions such as distributed denial of service attacks. This results in reducing the performance of the device along with "blacklisting" the network for hosting malicious attacks.

Cluley in [25] describes the vulnerability of IoT devices to Mirai Botnet and stresses the importance of changing one's IoT device's default password. Greene in [26] points to a huge DDoS attack on IoT devices such as cameras, lightbulbs, and thermostats by a botnet. The use of default or weak passwords in IoT devices makes them more susceptible to such attacks.

VII. Data at the Edge

A. IoT reference model

According to Cisco's IoT reference model in Figure 1 [27], the data is in motion in the lower layer of IoT. The dynamic data comes from the sensors and there exists a continuous communication of messages to actuators. Recent advancements in the IoT architecture has added more processing near the Edge of the IoT network. Edge pushes the intelligent processing capabilities closer to the network edge, which gives flexibility and makes the system much more responsive. There is a slight difference between Edge and Fog computing. Fog pushes the intelligence to the fog node, which resides in local area networks, close to the data. At this node, some of the information might transmit to the cloud. However, the edge node directly pushes the data to the "thing." In some cases, the key data is transmitted to the cloud for further analysis.

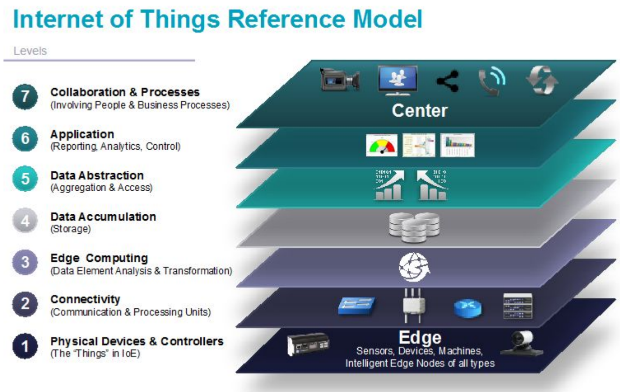

Fig. 1. Internet of Things(IoT) reference model [27]

B. Data acquisition

The sensors in Level 1 of the IoT Reference Model [27] are key sources of data in the IoT system. The sensors or "things" (such as computers) are connected to the Internet. IoT gateways provide an access route for devices without IP-address (such as lights, locks, gates, etc) to the Internet. A gateway provides a bridge between sensors, actuators and the Internet or Intranet with the use of different communication technologies. These communication technologies differ in terms of connectivity types, interfaces, or protocols. For example, IoT devices use more common technologies such as Bluetooth, LE, ZigBee, and Z-wave. Given the volume of the data collected by sensors, data filtration reduces the amount of data that is forwarded to the back-end for further processing or analysis. Also, edge computing helps to provide the IoT gateway security.

The IoT architecture connects devices directly to the cloud for processing and analysis. In Figure 2, all data from the sensor is sent to the cloud which leads to an unnecessary traffic and security risk. Waiting for messages to and from increases latency, which might affect real-time responses. This may not be favorable in emergency situations. It requires the resources to store and process the data, which is expensive. From Figure 2, we can estimate the latency of each part of network as:

$$Latency = T1 + T2 + T3 + T4$$

With a gateway, T2, T3, T4 are replaced by much faster interactions and data is transmitted to the cloud only when needed. This reduction in transmission of data requires considering the rate and type of data. Sensors in the IoT system collect huge amount and variety of data, which results in considering the combination of four V's in making a necessary decision. Cisco in [28] describes how devices send the right data to cloud for big data analytics and storage.

VIII. Data in the Cloud

A. IoT reference model

From the IoT reference model in Figure 1, the accumulated data in level 5 is abstracted for analysis. It involves processing with the queries on data sets. The data is first cleaned using

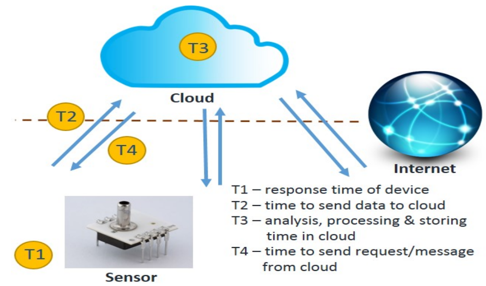

Fig. 2. Total time for an IoT response

various techniques such as normalization, standardisation, and other terminologies prior to the analysis and is then made available to level 6. Here, software applications of IoT devices provide back-end support for users. It generates business intelligence reports, analytics for decision-making, system management and other uses to control the IoT system. Level 7 involves collaboration and processes beyond the IoT network and application.

B. Data cleaning

Before the data collected by the sensor is ready for analysis, this raw data is required to be cleaned to make it technically correct and consistent. It should be done systematically and should be well documented for reproducibility and possible automation. Jonge et al. in [29] explain the steps involved in improving and refining data. The collected data comes with some identification. To reach technical correctness this raw data is encoded, decoded, converted, stripped, tagged and combined with meta-data. After this processing, data may still be inconsistent and unexpected. It requires domain knowledge of the IoT device to get past any compilation errors in the system. This processing is required before analysis in level 5.

C. Why is data stored in cloud?

Cloud infrastructure services such as Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS) allows organizations to avoid the need for in-house equipment, power, networking and IT support. Cloud, as a part of the Internet, can be accessed from anywhere, can shrink and grow according to the consumers demand. Clouds can be both public and private. Clouds such as Amazon Web Services (AWS) and Microsoft Azure, Google Cloud Platform are public clouds whereas private clouds sit within the security firewalls of an organization.

IoT devices have relatively small storage and processing power. The big-data generated from the IoT devices is stored, aggregated, processed and analysed in cloud. Moving the data towards the cloud gives "infinite" processing and storage capabilities. Below the cloud there are data centers, with numerous severs or host computers. Each host computer has multiple instances of Virtual Machine (VM) running as an application on the actual hardware, looking as a separate machine. The specification of these instances are taken into account and thus, the organisation pays for the additional resources used. VM's are an example of IaaS. However, web, blog-hosting, and IoT platforms are an example of SaaS which are more expensive than primary IaaS.

D. Cloud architecture

Figure 3 represents the IBM reference architecture infrastructure [30]. Cloud services such as IaaS, PaaS and SaaS are on the top left, while physical infrastructure is in the lower section. Consumer tools and in-house IT are used by users to interact with the Cloud Services. The service creation tools allows sustaining cloud resources along with important non-functional aspects such as security, performance, resilience, consumability, compliance and overall governance. Cloud architecture can be virtualized across many data-centers.

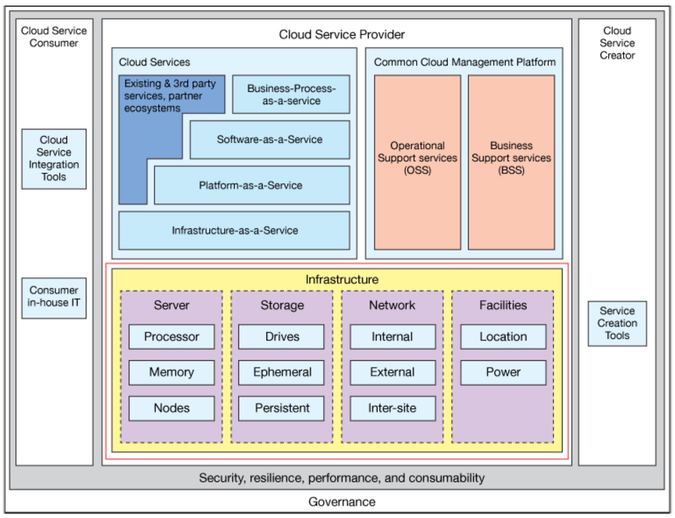

Fig. 3. IBM Cloud reference architecture [30]

E. Cloud based IoT platforms

IoT platforms include a dashboard to display and control devices. Additional features such as data collection, data management, testing, software updates and inventory management are also prominent.

Amazon Web Services (AWS) [31] is an IoT platform that includes a wide range of tools and services to deploy, setup and manage IoT solutions. It consists of four main products. They are:

- AWS IoT Core - base to built an IoT application
- AWS IoT Device - allows easy addition and organization of devices
- AWS IoT Analytics - provides service for automated analytics of massive amount of varied IoT data, including different data types
- AWS IoT Device Defender - support security mechanism of IoT systems

The AWS environment provides scalable and secure environment for IoT systems.

Google Cloud IoT [32] builds and manages IoT systems of any size and complexity. This cloud service includes:

- Cloud IoT core - allows connecting various devices and collects their data
- Cloud Pub/Sub - provides real-time stream analytics and processes event data
- Cloud Machine Learning Engine (ML) - allows the building of ML models and use of data received from IoT devices

Google Cloud IoT includes a number of service that might be useful for building a comprehensive connection of networks.

Microsoft Azure IoT Suite [33] provides security mechanisms, easy integration, and scalability. The Suite can easily connect to many devices from different manufacturers, collects data analytics and use the IoT data for machine learning purposes. The suite also offers preconfigured and customisable solutions to match requirements of the project.

Cisco IoT Cloud Connect [34] presents an end-to-end convenient platform for mobile cloud based IoT solutions. This service supports data and voice communication, customization of IoT applications and various monetization opportunities. The cloud consists of a complete package of monitoring functions, device management, advanced security measures, and scalability. With the growth of IoT devices, Cisco developed the kinetic platform supporting Edge and Fog computing. The kinetic platform manages IoT devices and gateways by giving support for data reduction, event processing, response, and data transfer to the cloud.

IX. IoT and Big Data Visualization

Big data generated by IoT devices (after collecting and analyzing) have to be represented in a visual way that allows humans to understand such analyses in an intuitive way. Visualization often allows gaining additional benefits or interpretations from a data set, providing more meaningful information. Along with this, presenting convincing graphics of the data helps to communicate those results to a wider range of audiences. Many algorithms and statistical methods are used on a large-scale and high-dimensional varied data which helps in the visualization of those data sets. The relationship between geometric objects within a data set is established using various parameters. Therefore, data visualization has become an important strategy for many business organizations to generate maximum revenue by improving decision making. There are several very powerful data visualization tools and techniques developed for IoT applications.

A. Data Visualization Techniques

Techniques such as simple plots, charts, maps, line or bar graphs, diagrams and matrices can be a very powerful way of highlighting any inconsistencies in the data set. This allows uncovering complex tables or numerical summaries and easy understanding of the results. Several techniques such as matrix methods in data mining, aggregations of attributes, dimensionality reduction techniques [35], [36] are highly used. Big data visualization cannot be approached using conventional techniques. Wang et. al in [37] propose a method called Discriminative Generalized Eigendecomposition (DGE), based on separation of multi-dimensional feature that could be useful in finding better discriminant vectors. This method deals with both Gaussian and non-Gaussian distribution. Zhong et. al in [38] proposed a RFID-Cuboid model which visualizes the real-time big data from cloud. This model can be used by end-users for their daily operations in a practical and feasible way.

B. Data visualization Tools

It is important to decide on the appropriate tool to be used for visualization to utilize the full potential of the collected data. Before exploring different visualization platform, the organization should identify its end-goals, identify its purpose, keep in mind its target audience, and should concentrate how to make the context more appealing. Several very sophisticated tools such as Plotly [39] and Sisense [40] provide some level of data analytics along with data visualization. Plotly enables the user to build charts using R or Python programming languages. It builds custom web applications using Python, provides access to open sources libraries for R, Python and JavaScript. Sisense is a cloud-based platform that has easy to use drag-and-drop interface. It supports natural language queries and can handle multiple data sources. Tableau [41] is a leading data visualization tool that has easy interface and interactive visualizations. Many large organizations rely on Tableau to generate meanings from their collected data. It has features such as automatic update, quick sharing, smart dashboards, and so on. There are several other tools that can deal with massive and complex IoT data. Microsoft Azure and Power BI [42], outstanding tools that can deal with any amount and type of real-time data. It provides several analytical power such as large integration capabilities, learning curve, along with drag-and-drop interface. ELK stack Kibana [43] is another tools that provide certain advance analytics such as exploring correlation between different observations, machines learning features to identify relationships between data events and so on. Grafana [44] provides services to query, visualize, create alerts and notifications along with several other capabilities.

X. Challenges and Future research

With the rate of data growth and expansion of IoT networks, it is important to have an accurate data of the environment. Organizations should acquire a specific skill set to deal with the analytical analysis of big data. The data collected by the organizations should be well structured and should be made compatible for use. To meet the demands of accurate data, it is necessary to connect a wide range of devices at any point and at any time. Therefore, there is need for investments in the field of sensors, data security and analytical capability to meet supply chain demands. The collection, processing, analysis and visualization of data set is a challenging task. Analysis of data based on specific data formats can limit the efficiency of the results. It is important to have full knowledge of the IoT domain in order to decide on the structure and format of the data collected by the sensors. Lack of this knowledge might result in dirty or garbage data, which can be costly. The issue of the 4 V's also pose a challenge while dealing with big data in IoT.

Nerkar et. al in [45] discuss data isolation in cloud computing as another challenge. Common resources shared in a cloud platform may cause the problem of inconsistency and latency in data content. Erway et. al in [46] describe about the challenge of efficiently proving the integrity of data stored at dishonest cloud servers. Patil et. al in [47] addresses security and privacy challenges as applied to the healthcare industry. As IoT devices collect and analyze data in a decentralized model, performing exhausting analysis operations while preserving privacy might be a challenge.

Even though the current technologies have achieved great results, there exists a wide scope in security and privacy concerns for the data collected by IoT devices. The communication overheads between the IoT devices that lead to latency must be optimized to achieve efficient results. With the growth of huge data, there is exist, storage overhead on the servers.

Consumers who use IoT devices for personal use might lack the technical knowledge required to understand or process the software requirements of the device. Some IoT devices and their software lack accurate information for users to make consenting decisions. It is necessary to make IoT software for personal use user-friendly and should always requests user's consent before sharing or making any decision.

XI. Conclusion

IoT has transformed many domains such as healthcare, infrastructures, manufacturing, retail, personal use and so on. As the data collected by IoT devices became big it became necessary to analyze this Big Data. Big Data has recently become more prominent in the IT technology, where it helps in product optimization, improves decision making and saves energy. As a result, Big Data has contributed substantially to IoT technology. Considering the huge amount of complex data produced by IoT devices, the analysis and visualization of that data has helped organizations meet demands and gain real-time business insights. Along with this, edge computing and cloud computing play highly important roles in aggregating large amounts of data and managing big data from anywhere in the world.

This paper does restrict itself to big data techniques in IoT but these techniques themselves are very viable for future research. In this paper, we discussed the issue of 4 V's in Big Data and how they are related to IoT. We discussed various data structure and data management approaches that should be used while managing Big Data in IoT. We discussed which layers in IoT reference model functions with respect to the existing and developing technologies such as edge and cloud computing. We also presented various cloud based IoT platforms, their key features and how they support organizations to handle massive and complicated big data. We discussed how data visualization approach is useful to interpret the meaning of data, along with several visualization tools and techniques. Lastly, we presented several challenges and future research work.

References

[1] Bernard Marr. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. *Forbes*, October 06 2017. [Accessed on: 06/01/2019].

[2] Kashmir Hill. *How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did*, volume Welcome to The Not-So Private Parts where technology privacy collide. Feb 16 2012. [Accessed on: 06/01/2019].

[3] DATA.GOV. *The home of the U.S. Governments open data*. [Accessed on: 06/01/2019].

[4] Amazon Web Services. *Registry of Open Data on AWS*. [Accessed on: 06/01/2019].

[5] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645 – 1660, 2013. [Accessed on: 06/01/2019].

[6] In Lee and Kyoochun Lee. The Internet of Things (IoT): Applications, investments, and challenges for enterprises. volume 58, pages 431 – 440, 2015. (Accessed on: 06/01/2019).

[7] H. Arasteh, V. Hosseinnezhad, V. Loia, A. Tommasetti, O. Troisi, M. Shafie-khah, and P. Siano. Iot-based smart cities: A survey. pages 1–6, June 2016. (Accessed on: 06/01/2019).

[8] Somayya Madakam, R Ramaswamy, and Siddharth Tripathi. Internet of Things (IoT): A literature review. *Journal of Computer and Communications*, 3(05):164, 2015. (Accessed on: 06/01/2019).

[9] Min Chen, Shiwen Mao, and Yunhao Liu. Big Data: A Survey. *Mobile Networks and Applications*, 19(2):171–209, Apr 2014. [Accessed on: 06/01/2019].

[10] J. Archenaa and E.A. Mary Anita. A Survey of Big Data Analytics in Healthcare and Government. *Procedia Computer Science*, 50:408 – 413, 2015. [Accessed on: 06/01/2019].

[11] Ulf Wetzker, Ingmar Splitt, Marco Zimmerling, Kay Rmer, and Carlo Alberto Boano. Troubleshooting Wireless Coexistence Problems in the Industrial Internet of Things. 08 2016. [Accessed on: 06/01/2019].

[12] Ishwarappa and J. Anuradha. *A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology*, volume 48. 2015. International Conference on Computer, Communication and Convergence (ICCC 2015). [Accessed on: 06/01/2019].

[13] Nawsher Khan, Mohammed Alsaqer, Habib Shah, Gran Badsha, Aftab Ahmad Abbasi, and Solmaz Salehian. The 10 Vs, Issues and Challenges of Big Data. pages 52–56, 03 2018. [Accessed on: 06/01/2019].

[14] Amir Gandomi and Murtaza Haider. *Beyond the hype: Big data concepts, methods, and analytics*, volume 35. 2015. [Accessed on: 06/01/2019].

[15] S. B. Baker, W. Xiang, and I. Atkinson. *Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities*, volume 5. 2017. [Accessed on: 06/01/2019].

[16] Mattia Antonini, Massimo Vecchio, Fabio Antonelli, Pietro Ducange, and Charith Perera. *Smart Audio Sensors in the Internet of Things Edge for Anomaly Detection*, volume PP. 10 2018. [Accessed on: 06/01/2019].

[17] Yoosang Park, Jongsun Choi, and Jaeyoung Choi. *Conceptual metadata model for sensor data abstraction in IoT environments*, volume 383. 07 2018. [Accessed on: 06/01/2019].

[18] Nicholas Dawes, K Ashwin Kumar, Sebastian Michel, Karl Aberer, and Michael Lehning. Sensor Metadata Management and Its Application in Collaborative Environmental Research. pages 143 – 150, 01 2009. [Accessed on: 06/01/2019].

[19] John P. Stevens. *Why you need metadata for Big Data success*. April 6 2016. [Accessed on: 06/01/2019].

[20] C. Sarkar, S. N. A. U. Nambi, R. V. Prasad, and A. Rahim. A scalable distributed architecture towards unifying IoT applications. In *2014 IEEE World Forum on Internet of Things (WF-IoT)*, pages 508–513, March 2014. [Accessed on: 06/01/2019].

[21] Sunil Kumar Mishra M. Handling the Unstructured Data in IOT. In *International Science Press*, pages 377–384, March 2016. [Accessed on: 06/01/2019].

[22] Mohammad Riyaz Belgaum Zainab Alansari, Safeeullah Soomro and Shahaboddin Shamshirb. A New Conceptual Model for BYOD Organizational Adoption. In *Asian Journal of Scientific Research*, volume 10, pages 400–405, 2017. [Accessed on: 06/01/2019].

[23] Serdar Yegulalp. What is NoSQL? NoSQL databases explained. In *InfoWorld, From IDG*, volume 10, Dec 7 2017. [Accessed on: 06/01/2019].

[24] Shuai Wang Xinyu Xing Lin Lin Heqing Huang Peng Liu Wenke Lee Le Guan, Jun Xu. From Physical to Cyber: Escalating Protection for Personalized Auto Insurance. pages 42–55, November 14-16 2016. [Accessed on: 06/01/2019].

[25] Graham Cluley. *These 60 dumb passwords can hijack over 500,000 IoT devices into the Mirai botnet*. October 10 2016. [Accessed on: 06/01/2019].

[26] Tim Greene. *Largest DDoS attack ever delivered by botnet of hijacked IoT devices*. September 23 2016. [Accessed on: 06/01/2019].

[27] Cisco Systems. The Internet of Things Reference Model. volume White Paper, 2014. [Accessed on: 06/01/2019].

[28] Cisco Systems. Cisco Fog Computing Solutions: Unleash the Power of the Internet of Things. volume White Paper, 2015. [Accessed on: 06/01/2019].

[29] Mark van der Loo Edwin de Jonge. *An introduction to data cleaning with R*. 2013. [Accessed on: 06/01/2019].

[30] Franck Barillaud, Chuck Calio and John A. Jacobson. *IBM cloud technologies: How they all fit together*. June 9 2015. [Accessed on: 06/01/2019].

[31] Amazon Web Services. *AWS IoT Core*. [Accessed on: 06/01/2019].

[32] Google Cloud IoT. *Google Cloud IoT"*. [Accessed on: 06/01/2019].

[33] Microsoft Azure. *Azure IoT Hub*. [Accessed on: 06/01/2019].

[34] Cisco IoT. *Internet of Things (IoT)*. [Accessed on: 06/01/2019].

[35] L. Eldn. *Matrix Methods in Data Mining and Pattern Recognition*. SIAM (2007). [Accessed on: 06/01/2019].

[36] D. Skillicorn. *Understanding Complex Datasets Data Mining with Matrix Decompositions*. Chapman Hall/CRC (2007). [Accessed on: 06/01/2019].

[37] Xiumei Wang, Weifang Liu, Jie Li, and Xinbo Gao. *A novel dimensionality reduction method with discriminative generalized eigen-decomposition*, volume 173. 2016.

[38] Ray Y. Zhong, Shulin Lan, Chen Xu, Qingyun Dai, and George Q. Huang. *Visualization of RFID-enabled shopfloor logistics Big Data in Cloud Manufacturing*, volume 84. Apr 2016. [Accessed on: 06/01/2019].

[39] Plotly. *Plotly*. [Accessed on: 06/01/2019].

[40] Sisense. *Sisense*. [Accessed on: 06/01/2019].

[41] Tableau. *Tableau*. [Accessed on: 06/01/2019].

[42] Microsoft. *Azure and Power BI*. [Accessed on: 06/01/2019].

[43] Elastic. *Kibana*. [Accessed on: 06/01/2019].

[44] Grafana Labs. *Grafana*. [Accessed on: 06/01/2019].

[45] R. R. Papalkar, P. R. Nerkar, and C. A. Dhote. Issues of concern in storage system of iot based big data. In *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC)*, pages 1–6, Aug 2017. [Accessed on: 06/01/2019].

[46] Chris Erway, Alptekin Küpçü, Charalampos Papamanthou, and Roberto Tamassia. Dynamic provable data possession. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*, CCS '09, pages 213–222, New York, NY, USA, 2009. ACM. [Accessed on: 06/01/2019].

[47] H. Kupwade Patil and R. Seshadri. Big Data Security and Privacy Issues in Healthcare. pages 762–765, June 2014. [Accessed on: 06/01/2019].