(IJIASE) 2019, Vol. No. 5, Jan-Dec **e-ISSN: 2454-9258, p-ISSN: 2454-809X**

EMPLOYABILITY OF THE INTERNET OF THINGS IN BIG DATA MANAGEMENT IN CONNECTED WORLD OF INTERNET OF THINGS A NOVEL APPROACH

Pavit

ABSTRACT

The Internet of Things (IoT) and Big Data have made the leap to turn into a mainstream issue and maintains board level priorities. Both IoT and big data are continuously creating headlines all over, drawing a huge amount of research interest and highlighting unique challenges. This growing recognition is due the intersection of both technologies with tremendous scope for business analytics and the prospective that still remains unexploited. Every day, industrial machine, health monitoring systems, sensors, and devices etc. connect to the Internet and exchange information. The future IoT will be extremely populated by huge quantity of heterogeneous networked embedded devices, which will be producing deluge of data. As businesses gets on new IoT project and tries to extract valuable information from enormous data volumes, a novel data management approach is called for. Conventional database management techniques and analytics methods fail to provide precise facilities to handle diverse data constantly flooding from numerous numbers of sources which are untold. This paper inspects the complex and fast moving data of IoT, and the existing position of data management techniques and challenges in storing and analysing it.

1. INTRODUCTION

Technological evolution has transformed the ways information was being accumulated, handled and evaluated by conventional processing systems. The whole process in today's era is executed automatically. With the merger of smart technologies (Radio Frequency Identification (RFID) [34] and wireless sensor networks (WSN)) and complex event processing CEP real time monitoring and communication with the physical world is possible. The improvement in networking techniques and information processing system have encouraged an advanced kind of an internet structure, called Internet of Things (IoT). The IoT refers to connected devices that can send and receive data over internet. The concept of IoT has been around for more than 15 years; however, it only began gaining extensive currency more recently2. Verizon defined the IoT as a machine to machine (M2M) technology based on cloud infrastructure with secure network connectivity3. Things in IoT must follow the three A's, autonomous (i.e. automatic data transfer to other devices or to Internet services), awareness (i.e. sense something) and actionable (i.e. integrate various sort of analysis or control). The IoT is an expansion of connectivity into a broader range beyond just machine-to-machine communication, which facilitates better data insights and analytics. Numerous amounts of data also known as Big Data is being generated by IoT. Big data analytic techniques are required to analyse of the flood of data from the physical resources to facilitate future Internet services. The objective is to comprehend the association between human and smart objects. The influence of IoT on human is still a big question and it demands consideration on how it plays a significant role in a smart world19. The future of internet technologies realizes in data and its analysis. Internet today is connected by objects and devices which are transmitting the assemble information for analysis. The objective is to utilize and find out the emerging trends in this data, which can have a positive impact on our society. The IoT revolutionizes the Internet by connecting physical world to the web through numerous sources resulting in deluge of data which requires proper process for collection, processing storage and analysis. To manage data generated effectively and intelligently extensive research and development is required which can provide information about our physical environment, at a level of detail never identified before1. Appropriate perception of accumulated data can bring about an enhanced understanding of the world we live in, building opportunities to improve way of living, working, learning and entertaining2. However, this disruptive technology requires new infrastructures due to dynamism of its networked participants as well as influx of heterogeneous data. Most advanced automated data analysis methods known today might fail to deal with the flood of data that starts flowing and growing by the minute. Several research papers discus various IoT data types and characteristic in the sight of database management3,5. Data centric and energy efficient IoT database management approaches and challenges are also reviewed in various research works4. Some surveys reveal Big Data analysis and related challenges in IoT and offers centralized cloud based solutions6,7. Whereas some acquires a things centric perception and argue for data analysis and compression before transmission of data to a cloud9-11. In addition, a research highlights decentralized data analysis as open issue concerning infrastructure for an applications of well-known data analysis algorithms in an IoT context12. Data gathered from various sources for analysis may possibly present a comprehensive view on the interactions and relationships between physical entities, facilitating the conversion of raw data and information into long term knowledge and perception2. This paper surveys the associated works, research challenges and present efforts for the management of data in IoT.

2. LAYERED VIEW OF IOT

The IoT by now facilitates billions of people by connecting smart things aimed to identify ubiquitous objects, data acquisition and information processing for everyone13. Thousands of smart connected devices bring new understandings to people throughout the world, reducing costs, sometimes by billions of dollars. IoT infrastructure can be represented as layered architecture comprised of four layers.

- Sensing Layers (Things) is used to collect information in order to link physical world with internet. It includes various devices for example sensors (infrared), readers (RFID), camera etc. The significant element which differentiates IoT from other networks is pervasive awareness. It is the sensing layer which enables the real time management of behaviour and properties of connected objects.

- Network Layer (Gateway) transmits the sensor data to middleware for data abstraction and processing. Control over ubiquitous objects is provided by Network layer which is an IP based internet, public/private network or a wireless network.

- Middle layer integrates several functions which are critical element in IoT infrastructure. It includes management and communication between devices, data processing and security. It offers interfaces for applications by abstracting the complexities of sensing and network layers.

- Application Layer provides end user domain oriented IoT applications.

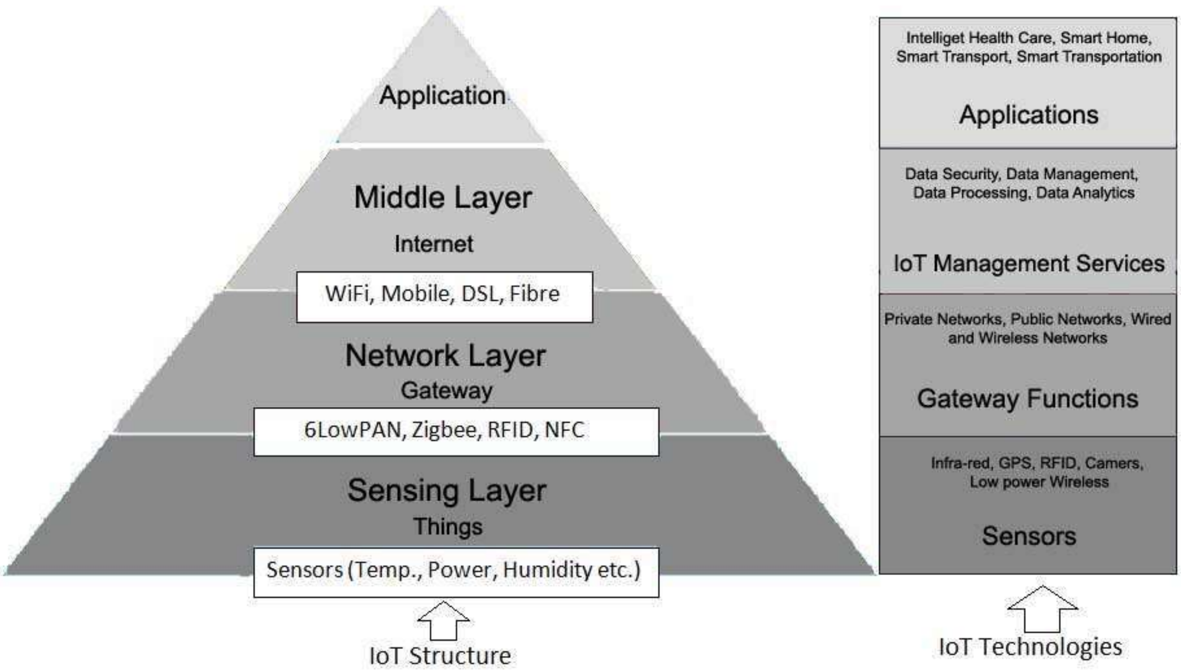

Figure 1. Layered View of IoT Data and Technologies.

The effective management and utilization of data in IoT layered architecture is a core reach challenge and needs attention. Data Management activity in IoT involves collection of data from heterogeneous sources which is then processed to convert raw data to useful information and finally stored for future analysis of data. Nevertheless, in general management structure and key topics relating data management and middleware needs close examination and extensive research.

3. DATA CHARACTERISTIC IN IOT

Day by day new devices, machines and sensors appear online and feed data into online storage systems. It is anticipated that by 2020 around 20.8 billion "things" will be operational globally15. IoT data is complex, enormous and fast moving, organizations attempt to extract more understanding from swelling data volumes and seek for new data management approaches to handle varied data continuously flooding from untold several of sources. Organizations that earlier derived their observations from transactional data are deviating towards IoT data. In a survey Aberdeen group examined IoT organization capability to accumulate, integrate, and analyse data generated by the disparate devices14. The investigation discovered the areas where organizations struggle and expect to progress for example the averaging volume of data grew 30% every year in IoT organization, with some organization complaining about insufficient data analysis capabilities and resulted fail to make timely decisions14. Embedded devices with sensing technology are more affordable than ever before and are connected over internet for continuously data transmission culminating data deluge which is also diversified. Current IoT organizations lack the analytical tools and infrastructure required to handle non-traditional data formats, such as geospatial and unstructured data. Policy makers are unable to convert this data into meaningful insights and are seeking for procedures which will allow them to process, store, and analyse this information. Secondly IoT organizations fail to make timely data driven decisions because they are unable to take rapid action on this fast streaming data. IoT main feature is real time or near real time communication of information about the connected things. IoT organizations require automated data analysis and management resolutions that yields rapid decisions, no matter how many end points are concerned. IoT is rapidly affecting the big data three characteristics (volume, variety, and velocity of data)36. The generation of intensive data or big data from heterogeneous devices will immensely confer the anticipated IoT. Usually the data mining techniques are used to extract knowledge from raw data20,21. The data collected by connected things have following characteristics:

- Data Variety, data is heterogeneous i.e. unstructured and semi-structured data for example as social media tweets, metadata, health records, audio / video streams, images etc. Organization face issue in performing data mining and machine learning analytics over the data to gain

- Data Volume and Velocity, Deluge of data due to real time exchange of data and information by heterogeneous things connected over a network. Efficient data filtering, compression and storage methods are required for this data processing.

- Data Inaccuracy is vital issue preventive the wide spread implementation of IoT. Sensing Technologies captures both reliable and an unreliable reading which further adds complexities in direct data usage.

- Data Semantics; the abstraction of complex semantics from the collection of unprocessed data with weak semantics in high level applications is required

4. IOT APPLICATIONS DOMAINS

IoT is an attractive element of every aspect of our lives. Its examples extend from smart connected cities to healthcare wearable34. IoT applications are intensifying the comforts of lives by controlling and simplifying routine work and personal tasks. The prospective of IoT world is enormous but some area that will develop much faster than the others. Some flourishing applications developed in fields like health care, smart environments, smart transportation, agriculture domain etc22,23.

4.1 IoT Applications in Connected/Smart Home and Smart Cities

In smart home the devices have ability to communicate with each other and with their intangible environment. There are several IoT technologies available for building and monitoring smart homes. For example, a smart home application can monitor home remotely, such as a control air conditioners and heater from remote devices as a tablet, phone or computer24,25. Smart cities IoT application includes environment monitoring, smart energy management systems, smart surveillance, safer and automated transportation.

4.2 IoT Applications in Wearable's and Healthcare

Wearable IoT technology is an extremely big domain broadly covering the fitness, health and entertainment requirements. The IoT wearable technology requirement is to be extremely energy efficient, low power and small sized. For example, wearable devices can sense the patient's medical data and sent remotely to his to pursue his health26.

4.3 IoT Applications in Automotive/ Transportation

In automotive and Transportation domain IoT offers various solutions for smart management. For example, smart parking can help drivers to save time and fuel by managing their car parking. It provides accurate information about available parking space which helps in reducing traffic jams24. Another example is of Google's self-driving cars.

4.4 IoT Applications in Agriculture

IoT application can deliver agriculture sector highly scalable technology solutions such as the Open IoT Pheno net Project which uses remote sensors to assist farmers to monitor air temperature, humidity and soil quality.

5. IOT DATA MANAGEMENT TECHNIQUES AND CHALLENGES

IoT development and adoption is rapidly accumulating the data, and experts warn that the present river of unstructured data will shortly convert into a flood. There are numerous methods and tools for solving several IoT data management challenges. A recent survey recommended concerns that most suggested techniques could lead to data management overload inadequate for the coming torrent of data. Conventional centralized databases will always play a part in analytics. However, IoT enterprises persist to gain momentum and moving from the central data repository towards the edge of the network. IoT organizations have automated data capturing process by embedding data management into the devices and sensors which are generating the data to expedite a smooth and stable stream of information. Therefore, data handled as soon as it is produced causing good control over real time data feeds. However, the data generated by IoT sensors or readers can be false data also known as dirty data in four forms False Positive (data in form of noise), False Negative (data loss in IoT sensing devices), Invalid and Redundant28. False Negative and False Positive data occur due to obstruction between sensors and environment, which may result in invalid data showing deviated values from the standard range. Data redundancy is caused because more than one sensor covers the same object. Data management process at this stage involves data cleaning, which eliminates false, incomplete, redundant and duplicated data, and solves the data quality issues in database system27. Techniques to clean data include spatial and/or temporal granule approach

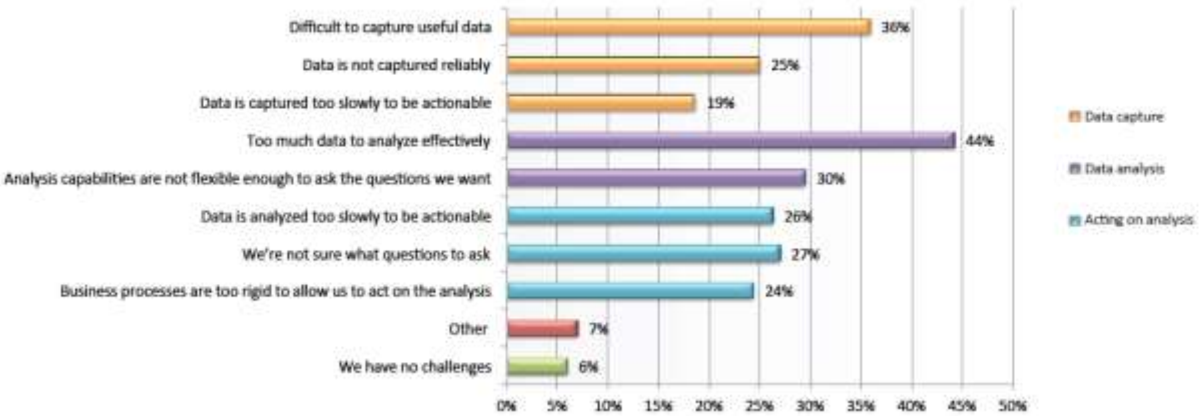

Figure 2. Challenges Faced in collecting and Analyzing IoT data.

and stream data cleaning framework. Well classified and filtered information is easier for analysis and eliminates congestion at central system and protects databases from overwhelming data volume and velocity. Organizations with IoT aspirations must invest in analytics at the edge to facilitate enhanced data management. Next stage is to process the accumulated cleaned data which can be treated as primitive event. Data deluge from heterogeneous sources with complex semantic is a big challenge for IoT data processing. Ontology based semantic event processing for IoT is an interesting research area and can be a significant tool to build the idea and association between "things" in IoT29. Once the relationship is established between IoT things, data can be exchange, storage, compressed and analyzed, that's where big data comes in; big data analytics tools are capable of managing loads of data transmitted from IoT devices that generate a nonstop torrent of information. The IoT brings data on a different range; the big data analytics solution must accommodate its requirements of speedy ingestion and processing followed by precise and rapid extraction. Two main computational paradigms for IoT big data processing are Map-Reduce and Data Stream30. MapReduce is a programming model, which distribute data to slave machines and perform computations in the sequence of map and reduce operations37. Whereas in data stream model algorithm process stream of data which is input in a sequence without explicitly storing it. Classical real time big data stream processing systems include the Twitter Storm31, LinkedIn Kafka33, and Yahoo's S432. Cloud computing has also become the typical platform for big data analytics. Technologies like SQream can deliver near real-time analytics on immense sized datasets, and effectively compress a full-rack database into a small server processing, therefore minimal hardware is required. The next generation analytics database leverages GPU technology, permitting further scale down of the hardware. This supports IoT organization to associate the increasing number of data sets in order to get real time response and get familiarized to the changing trends, overcoming the size challenge without negotiating on the performance.

6. KEY OUTCOMES OF CURRENT STATE OF IOT DATA

The present achievements of IoT organization analytics are still onset. Certainly the two field big data and IoT will create new possibilities that will have a long lasting impact. Enlightened IoT industries will persist to improve the processing, storage, and querying of IoT data and will consider confronts and techniques of data management for the IoT:

- It is expected that average organization's data will double within three years, therefore data is mounting and analytical demands are swelling. As the deluge of information intensify, decision makers calls for added capabilities and faster access. The current analysis capability of IoT organization analysis is insufficient and time-to-decision is not improving.
- Mature organization automatically filters and classifies data at the edge, ensuring relevant information and avoids overwhelming databases.
- IoT organizations considerably have analytical capabilities for unstructured and geospatial. Management of Data must be flexible enough to embrace diverse data types with integration of information in traditional formats.
- Right now, the majority of IT personnel at IoT organizations are not pleased with their ease-of-use of data systems. Organizations must think about automation schemes to speed up data processes and facilitate the user experience.

7. CONCLUSION

The Big data and IoT share a strongly knitted future. For the development of Internet of things, Big Data is a prerequisite. IoT is creating new business opportunities, improving customer experiences, accelerating growth and business performance, providing enhanced operation of machines and quality control, as well as improving safety. Without appropriate data assortment, businesses will be deprived of the opportunity to classify the information impending from the built-in sensors. Hence Big Data will be nothing more than white noise. The proper recognition of current patterns and trends in the data may perhaps further support proactive behaviour and planning, for example by foresee natural catastrophes, security breaches, traffic jams, etc. Nevertheless, rapid IoT data may expose early signs of lagging performance and the issue can be handled before it becomes a major problem that impacts customer experience. In the era of continuous technology evolution, a key component of any IoT application development the suitable database type is a fundamental element for ensuring success.