ADVANCEMENTS OF BIG DATA ANALYTICS IN IOT

[1]Akshitha, [2]Ramona Muthu, [3]Joyal K Scaria, [4]Roselita, [5]Precilla
[1, 2, 3, 4, 5] Student
[1, 2, 3, 4, 5] M.Sc. Big Data Analytics
[1, 2, 3, 4, 5] St Agnes Centre for PG Studies and Research, Mangalore, India

ABSTRACT:
The Internet of Things (IoT) and Big Data are two evolving technology topics in latest years. The Internet of Things (IoT) and Big Data go hand in hand. The main idea behind the IoT is that almost every object or device will be having an IP address and will be linked to each other. Currently, the term Industry 4.0 is used to describe the fourth industrial revolution that has enabled the digitization of the value chain. This revolution has also enabled the connection of production sites via intelligent information systems, which means that machines can communicate with other machines and products. In addition, more accurate data can be delivered and information can be processed in real time. The Internet of things (IoT) is the inter-networking of physical devices, vehicles (also referred to as "connected devices" and "smart devices"), buildings, and other items embedded with electronics, software, sensors and network connectivity which enable these objects to collect and exchange data.
Keywords*: *Industry 4.0, Big Data, Internet of Thing (IoT), Cloud Computing, Heterogeneous data, IoT architecture, IoT applications.

1. INTRODUCTION:

Big Data in IoT is a large and fast-developing area where many different methods and techniques can play a role. The IoT is a notion that depends on interconnected physical objects. It creates a mesh of devices that can able to generate information. IoT generates numerous amount of data therefore called "Big data," that provides advanced analytic techniques and offers a vision that makes machine usage easier and efficient. The Big data analysis is required to take advantage of its potential for high-level modeling and knowledge engineering. IoT enables us to know things that need replacing, repairing or recalling.

Since there is a massive growth in number of devices day by day, the amount of data generated would also be enormous. Here is where Big Data and IoT go hand in hand. Big Data manages the enormous amount of data generated using its technologies. The Internet of Things (IoT) and big data are two vital subjects in commercial, industrial, and many other applications. The Big data challenge is how to understand the interaction between human and smart objects. The basis of the Internet was human to human interaction when the human determines the content to be used by another human, but with the IoT the objects determine the content. Therefore, the impact on our lives is an open issue that needs understanding how IoT plays an important role in a smart environment and smart world.

The rest of this paper is organized as follows. Section II introduces to the General concepts of IoT. Section III discusses Importance of Industry 4.0. Section VI shows Big Data in IoT application areas. Section V discusses Techniques and Methodologies. Section VI shows IoT Architecture. Section VII discusses Impacts of IoT on Big Data. Finally, Challenges of IoT Big Data are discussed.

2. RESEARCH METHODOLOGY:

The paper is based on the secondary data and the information is retrieved from the internet via journals, research papers on the same subject matter.

2.1. General concept of the IOT

The Internet of Things (IoT) is one of the most rapidly emerging platforms for the digital economy. It is a web based network, which connects smart devices for communication, data transfer, monetary exchange and decision-making. Both the number of communication channels and the volume of data transmitted are increasing exponentially along with the number of devices that are connected to this network.

For the processing and analysis of very large data set "Big Data" a new research area and associated collection of methods and techniques have emerged in recent years. Although there is no clear definition for Big Data, a commonly quoted characterization are the "3V's": volume, variety, and velocity

Volume: There is more data than ever before. Its volume continues to grow faster than we can develop appropriate tools to process it.

Variety: There are many different and often incompatible types of data such as text data, sensor data, audio and video recordings, graphs, financial, and health data.

Velocity: Data can be streaming, that is, it is arriving continuously in real time and we are interested in obtaining useful information from it instantly. The ability to process depends not only on physical bandwidth and protocols but also on suitable architectural solutions and fast algorithms.

The IoT is regarded as the next phase in the evolution of the internet. It will enable commonplace devices to be connected to the internet to achieve many disparate goals. By the end of 2020, there could be up to 50 billion devices connected to the internet, far greater than the number of human users as shown in Figure 1 below. The growth in the IoT follows an exponential curve while the growth in the number of human users follows a logarithmic curve.

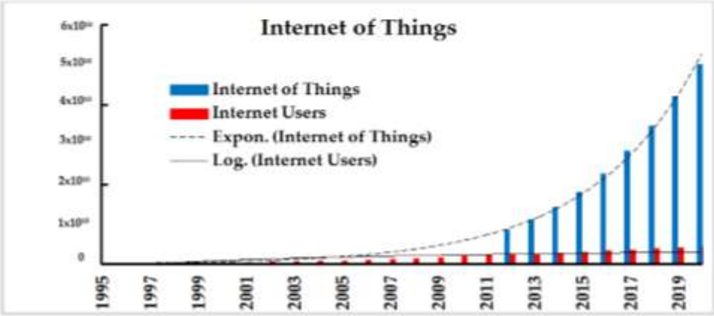

Figure 1. Internet of Things growth (data from http://www.internetlivestats.com/internet-users/ and IoT stats: statista.com). The red bars show the number of human internet users for the period 1995–2020; the blue bars show the number of devices connected to the internet, while the trend lines show a logarithmic growth for human users and exponential growth for number of devices connected.

The core components of the IoT will be sensors and actuators, embedded processing and connectivity and the cloud. Smart objects such as modern phones use sensors and actuators to interact with the real world. The IoT will ultimately evolve into a network of people, processes, data, and physical objects that intercommunicate using wireless protocols.

3. IMPORTANCE OF INDUSTRY 4.0

The reasons why industry 4.0 is important are the benefits. It helps manufacturers with current challenges by becoming more flexible and making reacting to changes in the market easier. It can increase the speed of innovation and is very consumer centred, leading to faster design processes. Workers can become coordinators at the centre of production, possibly improving the work-life balance of employees. Industry 4.0 is sustainable long-term, assisting in finding solutions for any challenge that arises.

Today, in an Industry 4.0 factory, machines are connected as a collaborative community. Such evolution requires the use of advance prediction tools, so that data can be systematically processed into information to explain uncertainties and thus make more informed decisions.

Power-generating technology such as robotization and automation has long existed. Industry 4.0 is the actual digitization of industry, which now covers a new, fairly broad conception, and includes new technologies and concepts relating to the organization of the value chain. Industry 4.0 creates a modularly structured smart factory, meaning the Cyber Physical System (CPS) monitors physical processes, maps the physical world in the virtual world, and decentralizes operational decision-making (autonomous machines).

Overall, we can conclude that Industry 4.0 penetrates the entire value chain of the corporation, although most of the value chains are interpreted as production-based, possibly supplemented with the logistics operations. The scope of Industry 4.0 can grow at the company’s borders, covering the supply chain or, more broadly, the supply network. It builds on new network-linked technology (e.g., sensors, RFID), and requires new procedures (e.g., data analysis software, cloud, programming) that require new capabilities from the company (e.g., continuous innovation, life-long learning, trust, data sharing) and this may even require new business models to be developed. Industry 4.0 is thus a phenomenon that, by means of technology assets and activities, maximizes the transparency of processes by exploiting the possibilities of digitization and integrates the corporate value chain and the supply chain into a new level of customer value creation.

4. BIG DATA IN IOT APPLICATION AREAS

4.1. Smart Environment Domain

- Smart home:

Stojkoska and Trivodaliev have outlined and proposed a generalized framework for an IoT-based smart home (Stojkoska & Trivodaliev, 2017). Their framework connects the home, utilities and third-party application providers through to a cloud network, with sensors attached to the smart grid system gathering data from smart home appliances.

Sensors, actuators, and controllers can be added to several home and office devices as a fan, fridge, washing machine, air conditioner and microwave. For example in Turkey, they apply an application for a home that is a solution for many problems. This application can monitor home remotely, detect fires, protect home from thefts and control devices as a heater and air conditioners from remote devices as a tablet, computer or phone.

- Smart environment control:

Many manufacturing enterprises have strict requirements on equipment working conditions and environment conditions for high-quality products. In the product manufacturing process, data on working condition variables and environmental conditions need to be gathered, stored and analyzed in real time to identify risks and abnormalities.

Environmental parameters such as temperature and humidity are important for agricultural production. Sensors are used by farmers in the field to measure such parameters and this data can be used for efficient production.

Environmental parameters measured in terms of temperature, soil information, and humidity are measured in real time and sent to a server for analysis. This data is saved and analyzed for extracting useful information such as the sample size, time, location and amount of residues. We can thus maintain the quality of the crop.

- Smart city:

Rapid growth of city populations due to urbanization has resulted in a steady increase in connectivity which in turn has generated a massive and heterogeneous amount of data. Increasingly, Big Data analytics is providing a better understanding of urban activities to support both current management and future planning and development.

Smart city is another powerful application of IoT generating curiosity among world's population. Smart surveillance, automated transportation, smarter energy management systems, water distribution, urban security and environmental monitoring all are examples of internet of things applications for smart cities. IoT will solve major problems faced by the people living in cities like pollution, traffic congestion and shortage of energy supplies etc. By installing sensors and using web applications, citizens can find free available parking slots across the city. Also, the sensors can detect meter tampering issues, general malfunctions and any installation issues in the electricity system.

- Smart Water Systems:

The smart cities can be able to detect water loss problem before it occurs and so it can significantly save on the budget. It helps the smart cities to detect the water leak sites and identify reform priority to prevent much amount of water from loss.

Given the prevailing amount of water scarcity in most parts of the world, it is very important to manage our water resources efficiently. As a result most cities are opting for smart solutions that place a lot of meters on water supply lines and storm drains. Smart water metering systems are also used in conjunction with data from weather satellites and river water sensors. They can also help us predict flooding.

- Wearables:

Wearable devices are installed with sensors and softwares which collect data and information about the users. This data is later pre-processed to extract essential insights about user. These devices broadly cover fitness, health and entertainment requirements. The pre-requisite from internet of things technology for wearable applications is to be highly energy efficient or ultra-low power and small sized. Wearables have experienced a explosive demand in markets all over the world. Companies like Google, Samsung have invested heavily in building such devices.

4.2. Transportation

Smart transport applications can manage daily traffic in cities using sensors and intelligent information processing systems. The main aim of intelligent transport systems is to minimize traffic congestion, ensure easy and hassle-free parking and avoid accidents by properly routing traffic and spotting drunk drivers. The sensor technologies governing these types of applications are GPS sensors for location, accelerometers for speed, gyroscopes for direction, RFIDs for vehicle identification, infrared sensors for counting passengers and vehicles and cameras for recording vehicle movement and traffic. There are many types of applications in this area:

- Smart parking: The smart parking offers solutions for management of parking that can help drivers to save time and fuel. By providing accurate information about vehicles parking spaces, it will be useful for making traffic flow better and reduce traffic jam.
- 3D Assisted driving: Vehicles like cars, buses and trains that are fitted with sensors can give useful information to the driver to save better navigation and safety. By using 3D assisted driving, the drivers can determine the correct path based on prior knowledge about traffic jam and accidents.
- Smart traffic lights: Traffic lights equipped with sensing, processing and communication capabilities are called smart traffic lights. These lights sense the traffic congestion at the intersection and the amount of traffic going each way. This information can be analyzed and then sent to neighboring traffic lights or a central controller.

4.3. Health care domain

- Health tracking:

Radio Frequency Identification (RFID) technology is useful for monitor person's health. The patient's medical data can be measured by sensing devices and sent remotely to his to pursue his health. IoT applications involve connecting the sensor to a person that can track the user's heart rate or pressure of blood continuously, for investigation via software or mobile applications

- Pharmaceutical products:

Smart pharmacy is a perfect application that helps easy accesses to remedy. Sensors attached devices can monitor the state of the drugs. In the case of finding expired drugs, it will prevent it from accesses to the patient.

5. IoT in agriculture

With the continuous increase in world's population, demand for food supply is extremely raised. Governments are helping farmers to use advanced techniques and research to increase food production. Smart farming is one of the fastest growing field in IoT. Farmers are using meaningful insights from the data to yield better return on investment. Sensing for soil moisture and nutrients, controlling water usage for plant growth and determining custom fertilizer are some simple uses of IoT.

4.4. Futuristics applications Domain:

- Robot Taxi:

Smart robot taxis in smart cities can treat with each other and provide services when requested by people. Robot taxis can treat easily with traffic congestion. It can move without the drive. It can avoid accident occurring. Using sensors and GPS, it can detect the position of people who request the robot taxi. In the case of stopping when sensors notified that actuators set off recharging batteries, it can make simple maintenance and clean the car.

- City information model:

The City Information Model (CIM) is depending on the notion that suggests all buildings is tracked by the government and allowed to the third party. Smart economy, smart people, smart devices, smart mobility, smart governance, etc. can simply interconnect with each other. Smart cities models should be integrated to improve performance and efficiency of the system.

5. TECHNIQUES AND METHODOLOGIES

Big Data Analytics and Tools:

There are many techniques or methodologies that can solve IoT data processing and analytics issues in many concepts.

5.1. Hadoop:

Big data can be collected and handled by Hadoop. Hadoop is proposed to parallelize data processing through computing nodes to hurry computations and hide latency. There are two main components for Hadoop: Hadoop Distributed File System (HDFS) and Map Reduce engine. HDFS stores enormous data constantly set and reproduce it to the user application at high bandwidth. MapReduce is a framework that is used for processing massive data sets in a distributed fashion through numerous machines.

5.2. MapReduce:

MapReduce was constructed as a broad programming paradigm. The Map Reduce includes two abilities from existing functional computer languages: map and reduce. The MapReduce framework gathers all sets with the common key from all records and joins them together. MapReduce is one of the new technologies, but it is just an algorithm, a technique for how to fit all the data.

5.3. HBase:

HBase is a database model inside the Hadoop framework that looks like the original system of Big Table. The HBase has a column that operates as the key and is the only index that can be used to get back the rows. The data in HBase is also saved as (key, value) sets, where the subject in the non-key columns can be represented by the values.

5.4. Hive:

Map Reduce jobs are difficult to track the characteristics of reusable code as some jobs are business particular some of the time. Hive may be thought as the necessary portion of Hadoop system and views at the top that principally is the organization for the data warehouse. Hive cannot treat with applications and transactions of the real time those are achieved online. The motivation behind it is a complicated technique.

5.5. Pig:

The Pig implementation designed within the Hadoop framework to offer additional database as functionality. Pig also provides a scripting language called PigLatin that offers all the common concepts of SQL, such as projections, joins, sorting, and grouping. The PigLatin language offers a higher extraction level to the MapReduce framework, as a query in PigLatin may be converted into a sequence of MapReduce tasks.

5.6. NoSQL:

It is an abbreviation to Not only SQL, and the most usual notion for non-relational databases. Various types of NoSQL databases, which are keyvalue pair document, column-oriented, and graph databases, that permit programmers to display the data suitable to the structure of their used applications. Google, Facebook, Amazon, and several other enterprises use NoSQL databases.

6. IOT ARCHITECTURE

The architecture of IoT has two perspectives. The first is the old perspective that considers the IoT architecture as three layers. The second is the new perspective that considers IoT as five layers.

a) The Three Layers Architecture:

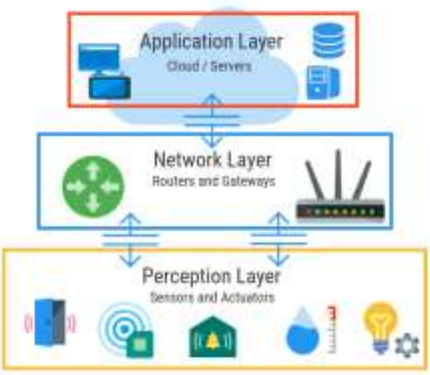

Figure. 2: The three layers architecture of the IoT

Fig. 2 shows the three layer architecture listed as follows:

- Network Layer:

This layer can transfer streaming sensor data from physical devices to the next layer. It can transmit the data to application devices via information processing system or infrastructure layer. The transmission channel may by wired or wireless, and the used technology can be 3G, Wi-Fi, Bluetooth, ZigBee etc. based on the sensor objects.

- Infrastructure Layer:

There are several elements of infrastructure that may result in many problems for users. Infrastructure should be promoted for services that are simply acceptable and to the expansion the velocity to the marketplace. Infrastructure commonly denotes to Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

- Application Layer:

The application layer is the great conjunction of IoT and industry technology, merged with industry requirements to recognize the industry, like person's social separation of work and the formation of the human economy.

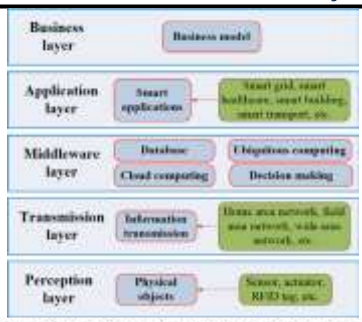

Figure. 3: The new IoT architecture (five layers)

b) The Five Layers Architecture:

Fig. 3 shows the five layer architecture listed as follows:

- Perception Layer:

It composed of physical objects and sensor devices. It is responsible for gathering physical object features such as humidity, motion, temperature, location and acceleration. It is done by varieties of sensors like RFID, 2D barcode and another sensor type. The collected data is transmitted from perception layer to network layer.

- Network Layer:

The network layer may be called as "transport layer." The medium of data transition may be wired or wireless and the using technology can be Wi-Fi, 3G, Zig-Bee, Z-Wave, etc. The basic task of this layer is transportation. It transports data from the network layer to the middleware layer.

- Middleware Layer:

Middleware layer is responsible for storing, analysing and processing the information of objects that received from the network layer and linked to the database.

- Application Layer:

This layer offered inclusive management of application that relies on the objects information that processed in the middleware layer. The implemented application of IoT may be transportation, logistics, smart health, smart home, smart city, etc.

- Business Layer:

Business layer likes a manager of IoT. The management includes applications, relevant system model and services. Technology success relies on technology priority as well as innovation of business model. The purpose of this layer is to determine the future actions and business strategies signals.

7. IMPACTS OF IOT ON BIG DATA

IoT is the next big thing impacting our lives in major ways and number of factors. Technologies like Column-oriented databases, SQL in Hadoop, Hive, Wibidata, PLATFORA, SkyTree, Storage Technologies, Schema-less databases or NoSQL databases, Streaming Big Data analytics, Big Data Lambda Architecture, Map-reduce, PIG, etc. helps in dealing with the enormous amount of data generated by IoT and other sources.

The main factors that big data is impacted by IoT are:

A. Big Data storage:

IoT has a direct impact on the storage infrastructure of big data. Collection of IoT Big Data is a challenging task because filtering redundant data is mandatorily required. After Collection, the data has to transfer over a network to a data center and maintained. Many companies started to use Platform as a Service (PaaS) to handle their infrastructure based on IT. It helps in developing and running web applications. By this way, Big data can be managed efficiently without the need of expanding their infrastructural facilities to some extent. IoT Big Data Storage is certainly a challenging task as the data grows in a faster rate than expected.

B. Data Security Issues:

The IoT has given new security challenges that cannot be controlled by traditional security methods. Facing IoT security issues require a shift. Few security problems are:

1. Secure computations in distributed environment
2. Secure data centers
3. Secure transactions
4. Secure filtering of redundant data
5. Scalable and secure data mining and analytics
6. Access control
7. Imposing real time security, etc.

A multi-layered security system and proper network system will help avoid attacks and keep them from scattering to other parts of the network. An IoT system should follow rigorous network access control policies and then allowed to connect.

8. CHALLENGES OF IOT BIG DATA

The challenges of handling Big Data are critical since the overall performance is directly proportional to the properties of the data management service. Analyzing or mining massive amounts of data generated from both IoT applications and existing IT systems to derive valuable information requires strong Big Data analytics skills, which could be challenging for many end-users in their application and interpretation. Some key challenges are listed in the following subsections:

8.1. Big Data mining:

Extracting values from Big Data with data mining methodologies using cloud computing now typically requires the following:

- Detecting and mining outliers and hidden patterns from Big Data with high velocity and volume.
- Mine geospatial and topological networks and relationships from the data of IoT.
- Developing a new class of scalable mining methods that embrace the storage and processing capacity of cloud.
- Providing new mining algorithms, tools, and software as services in the hybrid cloud service systems.

8.2. Architecture challenge:

In IoT, data collected from different resources so it will be difficult to integrate these heterogeneous data. The solution is collecting data from various sources and determining the common characteristics between them to explain data and find the associations for support decision-making. The existence of heterogeneous reference architectures in IoT is important. IoT architecture must be reliable and elastic to suit all cases as RFID, Tags, intelligent devices and smart hardware and software solutions.

8.3. Privacy and Security and Connectivity challenges:

- Privacy and Security :

Problems of security and privacy in IoT become more obvious than a traditional network. Todays, user's information has an extensive privacy, so privacy protection is a significant issue. Security architectures that are designed now may not be suited for IoT systems. Approval of new technologies and services depend on trustfulness of information and protection of data and its privacy.

- Connectivity challenges:

The future of IoT will very much have to depend on decentralizing IoT networks. Part of it can become possible by moving functionality to the edge, such as using fog computing models where smart devices such as IoT hubs take charge of timecritical operations and cloud servers take on data gathering and analytical responsibilities.

8.4. Data processing challenges:

Many data processing challenges are listed in the following subsections:

- Heterogeneous Data Processing:

In IoT applications, the enormous data are gathered from heterogeneous sensors such as cameras, vehicles, drivers, passengers and medical sensors. It results in heterogeneous sensing data like text, video and voice. For example, IoT system may contain many forms of sensors, such as traffic sensors, geological sensors and biomedical sensors. Each category can be separated into different forms of sensors. For example, traffic sensors can include GPS sensors, RFID readers, video-based traffic-flow analysis sensors, traffic loop sensors, road condition sensors and so on. The sampling data from different sensors may have dissimilar semantics and data structures that critically rises the troubles in data processing.

- Noisy Data:

Noisy data is irrelevant data. Statistical analysis can use the information collected from old data to clear noisy data and simplify data mining. Anomaly detection is the detection of irregular events or patterns that is not considered as expected events or patterns. Algorithms for anomaly detection are employed based on one type of learning formation: supervised, semi-supervised and unsupervised. The basic output of these techniques is qualified classifiers that accept new data as input and produce suggestion for data points as output.

9. CONCLUSION

This paper is a review about "Big Data Analytics role, in the field of IoT and Industry 4.0", which deals with the general concepts of IoT and the importance of Industry 4.0, where, IoT denotes to spreading the Internet of physical objects as a room, table or another human sensing objects as collections of features. They can be detected, determined and accessed by devices like actuators, sensors or other smart devices. Industry 4.0 penetrates the entire value chain of the corporation, although most of the value chains are interpreted as production-based, possibly supplemented with the logistics operations. The scope of Industry 4.0 can grow at the company's borders, covering the supply chain or more broadly, the supply network. Application of IoT has been presented in various fields and domains. Technologies related to big data have been surveyed from the perspective of data acquisition and network based. The different layers of IoT Architecture and its Impact have been discussed. Finally Challenges of IoT in Big Data have been discussed.

REFERENCES:

[1] Cesar Da Costa, Cleiton Rodrigues Mendes and Raphael Osaki – "Application of Big Data and the Internet of Things in Industry 4.0". EJERS, European Journal of Engineering Research and Science Vol. 3, No. 11, November 2018.

[2] Fayeem Aziz, Stephan K. Chalup and James Juniper – "Big Data in IoT Systems". booktitle = {Internet of Things (IoT): Systems and Applications}, year = {2019}, chapter = {2}, publisher = {Pan Stanford Publishing Pte. Ltd.}.

[3] Heba Aly, Mohammed Elmogy and Shereif Barakat – "Big Data on Internet of Things: Applications, Architecture, Technologies, Techniques, and Future Directions". International Journal of Computer Science Engineering (IJCSE), Vol. 4 No.06 Nov 2015.

[4] Pallavi Sethi and Smruti R. Sarangi – "Internet of Things: Architectures, Protocols, and Applications". Journal of Electrical and Computer Engineering Volume 2017, Article ID 9324035.

[5] K. R. Kundhavai and S. Sridevi – "IoT and Big Data- The Current and Future Technologies: A Review". IJCSMC, Vol. 5, Issue. 1, January 2016.

[6] M. Praveen Kumar, S. P Santhoshkumar, T. Gowdhaman and S. Syed Shajahaan – "A SURVEY ON IoT PERFORMANCES IN BIG DATA". IJCSMC, Vol. 6, Issue. 10, October 2017.

[7] Judit Nagy, Judit Oláh, Edina Erdei, Domicián Máté and József Popp – "The Role and Impact of Industry 4.0 and the Internet of Things on the Business Strategy of the Value Chain — The Case of Hungary". Sustainability 2018, 10, 3491; www.mdpi.com/journal/sustainability

[8] Peter J. Ryan and Richard B. Watson – "Research Challenges for the Internet of Things: What Role Can OR Play ?". Systems 2017, 5, 24; doi:10.3390; www.mdpi.com/journal/systems

[9] https://www.machinemetrics.com/blog/why-industry-4-0-is-important