THE ROLE OF IOT AND BIG DATA IN RENEWABLE ENERGY

Lappeenranta–Lahti University of Technology LUT

Bachelor's Programme in Electrical Engineering, Bachelor's thesis

Energy Economics

2024

Miro Kärki

Examiner: Dr. Mehar Ullah

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Energy Systems

Electrical Engineering

Miro Kärki

The role of IoT and Big Data in renewable energy

Bachelor’s thesis

2024

39 pages, 8 figures and 2 tables

Examiner: Dr. Mehar Ullah

Keywords: IoT, Internet of Things, Big Data, Renewable energy

Climate change is accelerating, and yet about 80% of the world's energy production is produced with fossil fuels. To get rid of fossil fuels, renewable energy production must be made as efficient and low-cost as possible. The purpose of this work is to find out how the Internet of Things and Big Data technologies can be utilised to improve the efficiency of renewable energy production and distribution, as well as to reduce costs.

This thesis has been carried out as a literature review. The information on the topic is sought with a preference mainly for English-language peer-reviewed scientific articles and other sources that can be found to be reliable, such as company websites. The thesis explores in more detail the concepts and operating principles of IoT, Big Data and renewable energy, as well as how these technologies interact with each other. Figures and tables have been used to facilitate understanding of various complex structures.

The study presented several ways in which IoT and Big Data can be utilised in the production and distribution of renewable energy to increase efficiency, as well as reduce costs. The methods found included remote monitoring and controlling of production plants, production and demand forecasting, and predictive maintenance based on IoT sensors. The work succeeded in compiling the potentially useful means on a general level, but further study will be needed in the future.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT Energiajärjestelmien Tiedekunta

Sähkötekniikka

Miro Kärki

IoT:n ja Big Datan rooli uusiutuvassa energiassa

Sähkötekniikan kandidaatintyö

2024

39 sivua, 8 kuvaa ja 2 taulukkoa

Tarkastaja: Tohtori Mehar Ullah

Avainsanat: IoT, Big Data, Uusiutuva energia, Älykkäät sähköverkot

Ilmastonmuutos on kiihtyvässä vauhdissa ja siitä huolimatta noin 80 % koko maailman energiantuotannosta tuotetaan fossiilisilla polttoaineilla. Fossiilisista polttoaineista eroon pääsemiseksi on uusiutuvan energian tuotannosta tehtävä mahdollisimman tehokasta ja matalakustanteista. Tämän työn tarkoituksena on selvittää miten Internet of Things ja Big Data teknologioita voidaan hyödyntää uusiutuvan energian tuotannon ja jakelun tehostamiseksi, sekä kustannusten laskemiseksi.

Tämä työ on toteutettu kirjallisuuskatsauksena, jonka pääaineistoina on käytetty vertaisarvioituja tieteellisiä artikkeleita, luotettavia verkkosivuja ja muita tieteellisiä julkaisuja. Työssä tutustutaan tarkemmin IoT:n, Big Datan ja uusiutuvan energian käsitteisiin, toimintaperiaatteisiin, sekä siihen miten nämä teknologiat toimivat vuorovaikutuksessa keskenään. Työssä on käytetty kuvia ja taulukoita erilaisten monimutkaisten rakenteiden ymmärtämisen helpottamiseksi.

Tutkimuksessa esiteltiin useita keinoja, joilla IoT:ta ja Big Dataa voidaan hyödyntää uusiutuvan energian tuotannossa ja jakelussa, sekä kustannusten alentamisessa. Löydettyjä keinoja olivat tuotantolaitosten etävalvonta ja -hallinta, tuotannon ja kysynnän ennustaminen, sekä IoT-antureihin perustuva ennakoiva kunnossapito. Työssä onnistuttiin kokoamaan mahdollisesti hyödynnettävät keinot yleisellä tasolla, mutta jatkotutkimuksia tullaan tarvitsemaan tulevaisuudessa.

ABBREVIATIONS

AI	Artificial Intelligence
HDFS	Hadoop Distributed File System
IEA	International Energy Agency
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
IPvX	Internet Protocol version X
LCOE	Levelized Cost of Energy
ML	Machine Learning
O&M	Operating & Maintenance
P2X	Power to Something
PV	Photovoltaic
RES	Renewable Energy Sources
RFID	Radio-frequency Identification
SCADA	Supervisory Control and Data Acquisition
SG	Smart Grid

Table of contents

1 Introduction

Climate change is one of the main challenges globally and it needs attention. It must be slowed down to ensure that the earth will remain habitable for our future generations as well. According to International Energy Institute (IEA, 2023a), the share of fossil fuels (coal, natural gas, and oil) in global energy production has been stuck at around 80% for decades. In addition to climate change, the use of fossil fuels has negative impacts for our environment and people's health (EESI, 2021).

In 2022, the world, particularly Europe and Asia, were plunged into an energy crisis. There are two main reasons for the crisis, first of all the extraordinarily rapid economic rebound after the pandemic but especially because Russia's invasion of Ukraine in February 2022 (IEA, 2023b). The consequences of war were immediately seen in the distribution of fossil fuels, especially on natural gas, which Europe's energy production is highly dependent. The huge increase in the price of natural gas led to energy prices reaching new all-time highs in 2022. Although, the situation at the electricity markets has already calmed down, people have started thinking more about their own electricity consumption behaviour, especially by reducing the use of electricity during peak hours. Electricity demand in European Union fell by 3% at 2022 compared to 2021 (Cam & Alvarez, 2023).

As the energy crisis and the negative effects of the use of fossil fuels show, the share of fossil fuels in global energy production should be reduced and they should be replaced with renewable energy. To improve the efficiency of renewable energy production and distribution we need to take all the advantage out of interesting new technologies such as Internet of Things, Big Data and Smart Grids to develop more sustainable energy solutions for energy producers and consumers.

1.1 Objective of the thesis and research questions

IoT and Big Data are already widely used technologies in many different fields, such as agriculture, smart cities, smart homes, health, transportation, and smart grids (Ullah et al., 2020). This thesis aims to increase understanding of how important role the Internet of Things and Big Data have in a field of renewable energy. The primary focus of this thesis will be how IoT and Big Data can improve efficiency of the renewable energy production and distribution. In this thesis, the concept of renewable energy is limited to solar photovoltaic (PV) and wind power only, as they are in this case the most significant renewable energy sources from the perspective of IoT utilization. The thesis also touches on how utilizing IoT and Big Data helps to reduce costs.

Research questions:

- *How IoT and Big Data improve efficiency of the renewable energy production and distribution?*
- *How IoT and Big Data reduces costs in the renewable energy production and distribution?*

Through this thesis, the aim is to map the current and future opportunities of Internet of Things and Big Data in the renewable energy sector. The study is expected to add value for renewable energy producers and consumers by providing a comprehensive overview of the current state and potential of IoT and Big Data applications in the field of renewable energy. The discovered benefits of utilizing IoT and Big Data in renewable energy will require further research in the future.

1.2 Research methods and structure

This thesis will be done as a literature review. The information on the topic is sought with a preference mainly for English-language peer-reviewed sources and other sources that can be found to be reliable. For example, various articles, websites, studies, or literature are used as sources to map the current situation as well as future prospects for utilizing IoT and Big Data in renewable energy sector. The structure of the thesis could be described briefly: Section 2 will explain IoT and the technologies behind it. Section 3 focuses on Big Data and its benefits. Section 4 describes renewable energy sources and their importance in energy production. Section 5 is explaining the relationship between IoT and Big Data and the requirements for adapting Big Data in IoT. The section 6 discusses about how IoT and Big Data can be used to improve the production and distribution of renewable energy. And finally, in the section 7, conclusions, where the results of the study are critically analysed and considered if there is a need for further research.

2 Internet of Things (IoT)

Internet of Things, from now on IoT, can be defined as “A network of physical objects or ‘things’ that can interact with each other to share information and take action” (Mandal, 2020). These physical objects or ’things’ can be anything, such as mobile phones, buildings, domestic appliances, sensors, vehicles, or lights. Basically, anything that can be connected to the internet. IoT is a global network of machines and devices which can interact with each other and with the platform. Devices and machines use sensors, actuators, communication, and analytical tools to generate data for the network to collect, analyze, store, and visualize (Ullah et al., 2021). The term ‘Internet of Things’ is perceived as being created by Kevin Ashton in 1999. He used it as a title of his presentation at Procter & Gamble in 1999. The concept of IoT began to take shape from RFID (radio-frequency identification) technology, which can be used to identify, trace the location and status of any physical object (Greer et al., 2019). As IoT has grown tremendously in popularity over the years, several other more advanced technologies have emerged alongside the simple RFID technology. IoT is an important part of the daily lives of many of us, and it has established itself as part of our modern society. It is considered as a unique and revolutionary technology of the future that has received a lot of attention in various industries such as in renewable energy which this thesis explores in more depth.

2.1 Key Components and Technologies of IoT

The main components or “building blocks” of IoT uses a wide range of technologies to enable the operation of the network. The objective of this section is to introduce the main components and technologies of IoT to understand its working principles and how the technologies can be utilized on the field of renewable energy.

2.1.1 IoT Building Blocks

IoT consists of six main components that work together and utilize different technologies to make IoT function properly (Ullah et al., 2020). These main components are called IoT building blocks, which reflects the functionalities of IoT.

1. *Identification block*. All devices in the network must be identified. Each device in the communication network has its own object ID which includes the name of the device and the object address in the network for the identification. Internet protocol versions 4 and 6 (IPv4 & 6) are the most used addressing methods of IoT objects. (Ullah et al., 2020)
2. *Sensing block.* IoT devices use many different types of sensors which range from very simple to advanced ones. The sensors are used for detecting the changes in the state of a physical object or in environment in order to obtain data that can be sent either to the other devices in the network or to the cloud, where the data can be analyzed. (Sehrawat D. & Gill N. S., 2019)
3. *Communication block.* The communication block is responsible for transferring the data collected by the sensors. It uses different communication protocols such as, CoAP and MQTT for connecting IoT objects and to send collected data to the management system. The sensors and IoT devices are connected to the internet by using communication technologies such as, ZigBee, NFC, UWB, Wi-Fi, SigFox, and BLE. (Ullah et al., 2020; Hejazi et al., 2018)
4. *Computation block.* The computation block represents the processing unit and provides the computational capability of IoT. It can be separated to two parts, hardware, and software. This means that the computations can be performed either on the hardware itself or through a software. There are many hardware platforms, such as Intel Galileo, Raspberry PI, Gargeteer, UDOO, and Arduino, which can be used running IoT applications. However, nowadays IoT platforms are typically cloud-based. (Ullah et al., 2020; Hejazi et al., 2018)

5. *Services block.* IoT services enable IoT application developers to focus on building the application instead of designing the service or architecture for the IoT application. IoT services are divided into identity-related, information aggregation, collaborative-aware and to ubiquitous services. Identity-related services can be split into two parts. Active identity-related services have a constant power, or a battery and they can send information to another device on their own, while passive identity-related services have no power source and they need an external device for identifying and they are only able to read information from the other devices. Collaborative services make decisions based on the data provided by information aggregation services, which collect and process the data and transfers it to the IoT application. Ubiquitous services can provide collaborative-aware services to anyone, regardless of time or place. (Ullah et al., 2020)

6. *Semantic block.* IoT is very complex system, because it includes a huge amount of heterogenous devices, sensing data and services. The purpose of semantic block is to act as the brain of the IoT. The semantic block facilitates effective data access and integration, resource discovery, semantic reasoning, and knowledge extraction. It finds and uses resources, models information, and recognizes and analyzes data to make decisions and to provide the correct service. (Ullah et al., 2020; Rhayem et al., 2020)

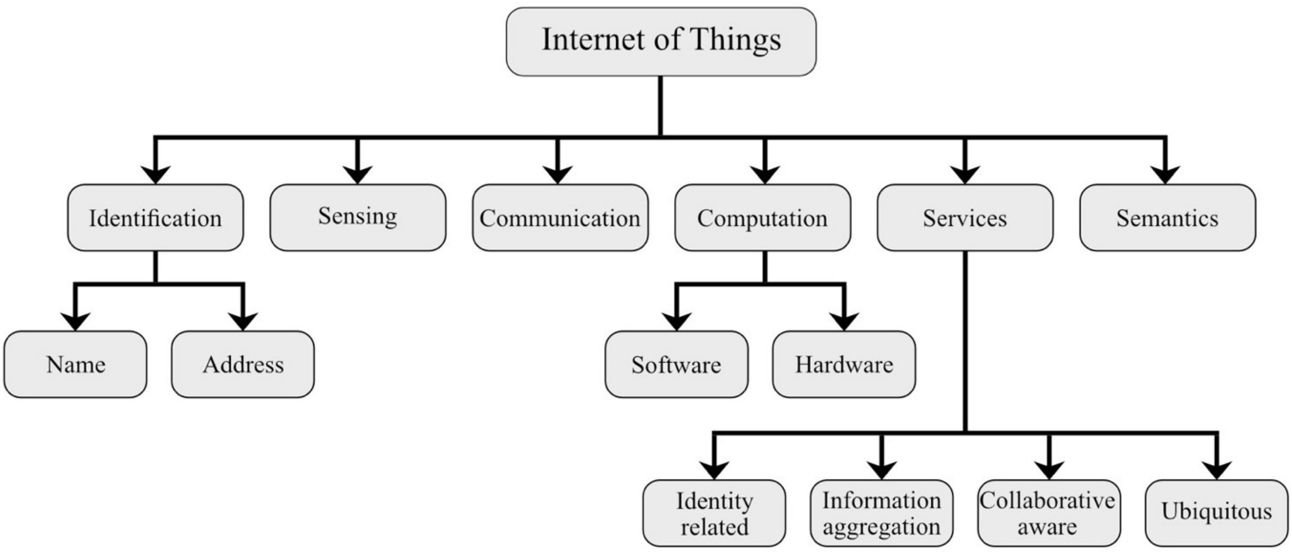

Figure 1. IoT Building blocks. Adapted from Ullah et al. (2020).

2.1.2 Cloud Computing

IoT devices and sensors continuously generate huge amounts of new data which needs to be stored and analyzed. Cloud computing moves data storage and analyzation away from devices into large data centers or into “cloud” (Jadeja & Modi, 2012). According to Ullah et al. (2021) many industries have shifted their business operations to cloud computing, because it provides flexibility, operability and cost saving in the operations. Cloud computing is especially good when a lot of computing power is needed, for example, in complex and intelligent big data analytics tasks (Dong et al., 2020). The most used IoT platforms offering cloud computing capabilities are Amazon Web Services (AWSs), Microsoft Azure, IBM Watson IoT Platform and Google Cloud IoT Platform (Ullah et al., 2021). These cloud platforms are used to store and process data to provide information for the end user based on the collected data.

Cloud computing is a good way to facilitate data storage and processing, but the continuously increasing amount of data can cause significant problems in cloud computing operations. First, most of the huge amounts of data sent to the cloud may be irrelevant for the operations performed, which means that the excess data ends up creating an extra load to a central repository and it increases costs due to the additional storage needed for the irrelevant data. Also, if the central repository is under a high load, it can increase the latency considerably which can be crucial in some operations. Another major problem is that sending and retrieving data from the cloud can be very expensive. Despite the bandwidth, latency, disaster recovery and cost problems with cloud computing, centralized data storage and analysis still plays an important role in modern data analytics. Especially data that needs a lot of processing power and isn’t time-sensitive can be sent to cloud for in-depth analysis. (Ullah et al., 2021)

2.1.3 Edge Computing

Edge computing is used reduce issues associated with cloud computing. It moves data analyzation away from the cloud to the “edge” of the network. More and more data is generated at the edge of the network, which would also make its analyzation more efficient at its place of origin. Edge can be defined to be any device between the cloud and the data collection device. From a cloud perspective, edge data is downstream data and upstream data behalf IoT services. In edge computing most of the data analyzation is performed at the source of data collection where the important data gets filtered and sent to the cloud for further analysis if necessary. This way edge computing can reduce response times, make bandwidth cost savings, as well as improve data safety and privacy. (Shi et al., 2016; Ullah et al., 2021)

According to Ullah et al. (2021), The network edge must be well-designed to meet privacy, security, and reliability requirements to perform tasks such as data processing, caching, device management, and privacy protection. By performing these tasks, edge computing can fulfill its purpose by reducing traffic between devices and the cloud with distributed data processing. Edge computing is the best in situations where important data needs to be processed quickly with a low latency and for example, if the processed data is controversial, it may not be wanted to be sent to the cloud for processing for security reasons. However, data processing latency can be affected by distance between data collection point and the edge processing server, edge server’s processing power, task’s complexity, and traffic on the edge. (Ullah et al., 2021)

In summary, collaboration between edge computing and cloud computing improves response times and reduces costs by reducing the traffic of irrelevant data between devices and the cloud. Edge computing makes distributed real-time data analysis possible in the remote areas and reduces bottlenecking on the networks and datacenters that support edge devices, while reducing the costs of data processing. (Microsoft, 2023)

2.2 How the IoT works?

The purpose of this section is to explain how IoT works using the scenario in Figure 2 as an example, which illustrates the layers of IoT in a renewable energy application. IoT in renewable energy consists of 6 layers starting from the upper left corner are: renewable energy sources (RESs), sensing layer, edge computing layer, network layer, middleware layer and application layer. All these layers are involved in monitoring and controlling RESs.

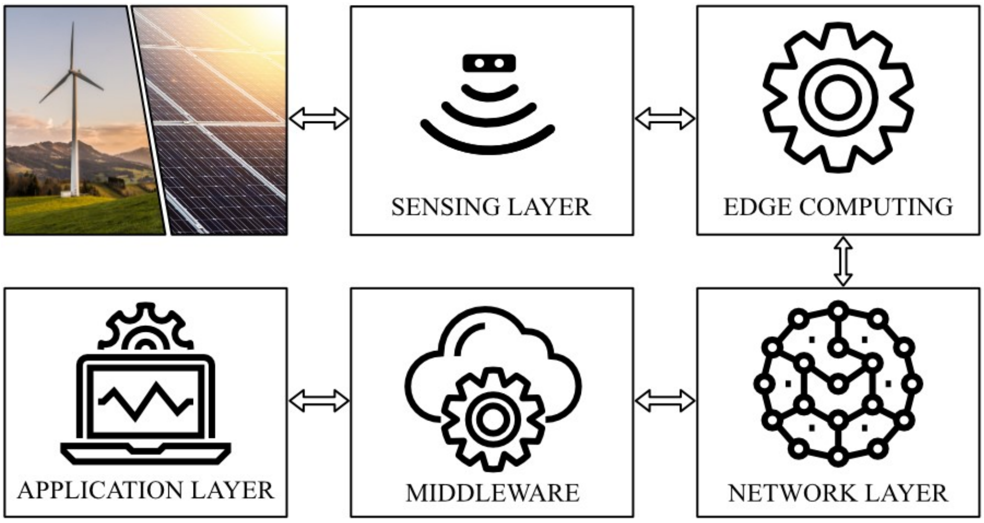

Figure 2. IoT layers for renewable energy, edited from (Ponnalagarsamy et al., 2021)

The first layer represents renewable energy sources, which in this thesis are limited to solar and wind energy only. The sensing layer is responsible for collecting data from the renewable energy sources and their surroundings. It uses various sensor devices and other components to collect detailed information about physical factors at the renewable energy plant. In the renewable energy, IoT sensors are used to monitor and manage the generation, transmission, and distribution of renewable energy remotely without human intervention. Sensors used in the solar and wind power production can be for example various electricity production, temperature, radiation, vibration, or wind speed sensors. (Sehrawat & Gill, 2019; Ponnalagarsamy et al., 2021)

The data collected by the sensing layer must be processed at some point, either in the cloud or at the edge to obtain useful information from the data. Data that does not need a huge amount of processing power can already be partially analyzed at the data collection point, which reduces the amount of irrelevant data sent to the cloud, thereby reducing costs, as well as additional traffic in the cloud (Ullah et al., 2021). Edge computing is also good in situations where you do not want to send data to the cloud for analysis, for example for security reasons.

The next layer is network layer which is responsible for information and data transfers through access and transport network. Access network is a short-range wireless network using technologies such as, Wi-Fi, Zigbee and NFC to connect the sensors and IoT devices to the internet. Transport network includes both, wired and wireless area network, and it uses technologies such as, IPv4 and IPv6 for the transport network access. (Ponnalagarsamy et al., 2021)

The middleware layer is a software layer between software applications that facilitates software developers' communication and input/output. It facilitates the work of IoT developers by hiding complex technologies behind software services that are irrelevant for the specific IoT application. The middleware layer simplifies a complex system and extracts the information and converts it into a required format. Most IoT middleware architectures follow service-oriented approach. (Ponnalagarsamy et al., 2021; Lee & Lee, 2015)

Application layer consists of the cloud computing and application platforms. The application layer stores, organizes, and analyzes data received from the sensing layer to make decisions and provide information to the user. On the platform the user can view visual graphs and reports generated from the data. Popular IoT platforms are for example Amazon Web Services (AWSs), Microsoft Azure and Google Cloud IoT platform (Ullah et al., 2020). In renewable energy, application layer can be responsible for functions such as, plant monitoring, fault monitoring, data analytics, maintenance, historical data analysis, generation monitoring and inspection scheduling. (Ponnalagarsamy et al., 2021)

2.3 Challenges of IoT

IoT is very useful technology, and it has many different applications on various fields, but there are also some challenges in its implementation. According to Lee & Lee (2015) IoT has five major technical and managerial challenges: data management, data mining, privacy, security, and chaos challenge.

The data management challenge is related to the huge amounts of data produced by the IoT sensors and devices that needs to be stored and analyzed. The huge amount of data needs a lot of storage and processing power from the data centers which can lead to increased latencies. To improve processing efficiency and response times, the data centers need to be more built more distributed in the future. The collected data is becoming more and more complex, which requires advanced data mining/processing tools. Traditional data mining techniques may not be able to process unstructured data, which can make it challenging to manage, store and analyze data effectively. (Lee & Lee, 2015)

The most alarming challenge in IoT is its privacy and security challenges. Nowadays, many consumers use wearable IoT devices in their daily lives. These devices may collect sensitive information about the consumer, such as location or health condition data, which is a significant privacy concern. Also, many IoT devices are poorly protected against data breaches, which can lead to compromising the network. Security is especially important in sensitive infrastructures and strategic services such as the smart grid and facility protection. However, safety problems can be solved by paying more attention to the proper protection of the IoT devices. (Lee & Lee, 2015)

Chaos challenge refers to the rapid and uncareful development of IoT innovations which may not have been tested or designed properly. These poorly designed devices can cause chaos in a world where all the devices are connected to each other. Poorly designed devices can have fatal consequences if they fail, for example in healthcare, or cause a disastrous chain reaction to the other connected devices. (Lee & Lee, 2015)

3 Big Data

In our technology-filled society, the amount and availability of data is growing at an extremely fast pace. Big data refers to these extremely large collections of unstructured, structured, and semi-structured data generated by different data sources, such as devices, sensors, software applications, and digital devices that are continuously generating data (Ullah et al., 2021). The amount and complexity of data is so enormous that the traditional data processing systems are not able to store, process and analyze them (Google Cloud, 2023).

Big Data is often defined through its three characteristics, the 3 Vs of Big Data, which are volume, velocity, and variety. *Volume* stands for the huge amount of continuously produced and collected data from various sources, which is making the datasets too large for the traditional data storage. *Velocity* of big data refers to the speed at which data is produced, processed, accessed, and analyzed. Nowadays, the whole process is done in real time or at least very near real time. *Variety* means that the data collected from many different sources is heterogenous, and may include structured, unstructured, or semi-structured data. Different data types are explained in the Table 1. (Ullah et al., 2021; Google Cloud, 2023)

Table 1. Different data types

Aspect	Structured	Unstructured	Semi-structured
Organization	Defined data model and highly organized	No defined form or specific format	Some organizational properties
Machine Readability	Easy to read	Hard to read	Moderately readable
Examples	Databases and spreadsheets	Documents, images, audio etc.	HTML, XML, log files, NoSQL databases

Big data is different compared to the traditional data processing, because the processed data masses are huge, the whole process must be done at high speed and the data can be significantly more complex than traditional structured data.

3.1 How the Big Data works?

According to Marjani (2017), Big Data analytics aims to extract information from the data to facilitate making predictions, identifying trends, finding hidden information, and most importantly making decisions. To achieve this, data must be collected, stored, and analyzed using various technologies. There are many different technologies and methods that can be used in Big Data analyzation, but in order to keep the thesis concise, we present only the most important ones.

Various sensors and devices continuously produce a huge amount of structured, unstructured, and semi-structured data. These huge amounts of complex heterogenous data that edge nodes cannot process are sent directly to the low-cost storage in the cloud. In the next phase of data acquisition, the data will be stored in a shared distributed fault tolerant database. (Ullah et al., 2021)

Since the collected data is generated from several different devices that may use different data formats, filtering out inaccurate and irrelevant data with data preprocessing will become necessary. Apache Flume is a distributed service, which main function is efficiently collecting, aggregating, and transferring the large amount of data to Hadoop master node. The data collected is written in a desired format using plug-in serializers in external HDFS (Hadoop Distributed File System) repository. After the Flume data is restructured into desired format the data preprocessing is ready and the data can be sent forward for processing. (Apache, 2024a; Ullah et al., 2021)

In the next phase, data will be transferred into the master node(s) in the Hadoop cluster, which is a collection of computers that are connected to each other to perform parallel computations of big data sets. Apache Hadoop is an open-source software that allows distributed storage and processing of large data sets across clusters of computers using simple programming models (Apache, 2023). Many big companies such as SwiftKey, Nokia, redBus and Alacer utilize Hadoop technology in different fields (Yaqoob et al., 2016).

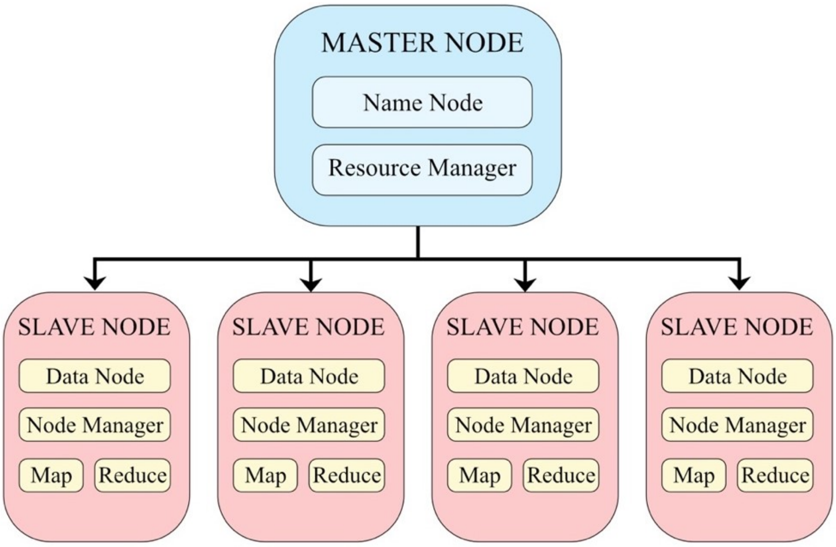

Figure 3. Hadoop Architecture. Adapted from Tyagi et al. (2022).

Figure 3 represents Hadoops master-slave architecture. To process huge amount of data, Hadoop uses Map/Reduce programming model. Master node divides a task into smaller parts and sends it to the slave nodes. After completing the tasks master node collects the information from the slave nodes and combines the small parts together to provide output. This data analysis process is performed by the YARN in the HDFS. Hive and Impala are the tools used to perform SQL queries on data. Hive enables summarization, querying and analysis of data. The final phase, data analysis, takes place in Hadoop using Scalable Advanced Massive Online Analysis (SAMOA), which offers distributed streaming algorithms for big data mining and machine learning tasks. Data visualization is implemented with Tableau, which is a common tool for live data visualization and interactive charts/graphs. (Yaqoob et al., 2016; Ullah et al., 2021; Bifet, 2015; Apache, 2024b)

According to Tableau (2023), Big Data analysis methods include data mining, predictive analytics, and deep learning. Data mining is a process where computers sort through large datasets to identify patterns and relationships by identifying anomalies and creating data clusters. Predictive analytics utilizes historical data to create trends and make predictions about the behavior in data. Deep learning uses artificial intelligence (AI) and machine learning (ML) to find pattern from abstract data. (Tableau, 2023)

3.2 Benefits of Big Data

Big Data analytics plays an important role in modern data analytics. As the amount and production of data grows, efficient big data processing becomes increasingly important. Nowadays, Big Data analytics has several significant technological and business benefits on many different fields.

From a technological point of view, it is extremely important that huge amounts of complex heterogenous data can be stored and analyzed efficiently. Big Data analytics can be used to process data that would not be possible with traditional data processing methods. Big Data also enables data processing and analysis in real time, which is very useful in multiple applications. Big Data analytics can be used to identify signs that a device needs maintenance, which can reduce device maintenance costs significantly (McKinsey, 2015).

From a business perspective, Big Data helps to reduce costs and, for example, improve customer experience. By utilizing Big Data analytics, companies save money, please customers, make better products, better decisions, and innovate business operations. Big Data analytics helps companies to extract information from complex data, which allows to make better business decisions and make predictions about the future. (SAP, 2023)

Table 2. The benefits of adopting Big Data technologies for companies (Raguseo, 2018)

Benefits	Transactional	Strategic	Transformational	Informational
1.	Productivity growth	Better products and service	Development of new business opportunities	Data management
2.	Reduction of operational costs	Aligning IT with business strategy	Accurate weather data	Data accuracy
2.	Improved return on financial assets	Faster response to changes	Improvement in employees' skills	Easier access to data

In her article, Raguseo (2018) have listed types of benefits companies can achieve by adopting Big Data technologies. According to Raguseo (2018), the benefits of adopting Big Data technologies for companies fall into four different categories: transactional, strategic, transformational, and informational. The three most recognized benefits for each category are listed in Table 2.

4 Renewable Energy

Renewable energy sources such as solar, wind, hydropower, biofuels etc. play an important role in the transition towards less carbon-intensive and more sustainable energy systems (IEA, 2024a). This thesis focuses only on solar and wind energy, as they are the most significant sources of renewable energy where IoT and Big Data should be utilized. The shift towards the wider use of renewable energy sources in energy production has received a lot of political support and more than 140 countries including China, US, India, and EU have set target to be net-zero by 2050 (UN, 2024). According to IEA (2024a) the renewable energy generation capacity has grown rapidly in recent years due to policy support and cost reductions for solar PV (photovoltaics) and wind power.

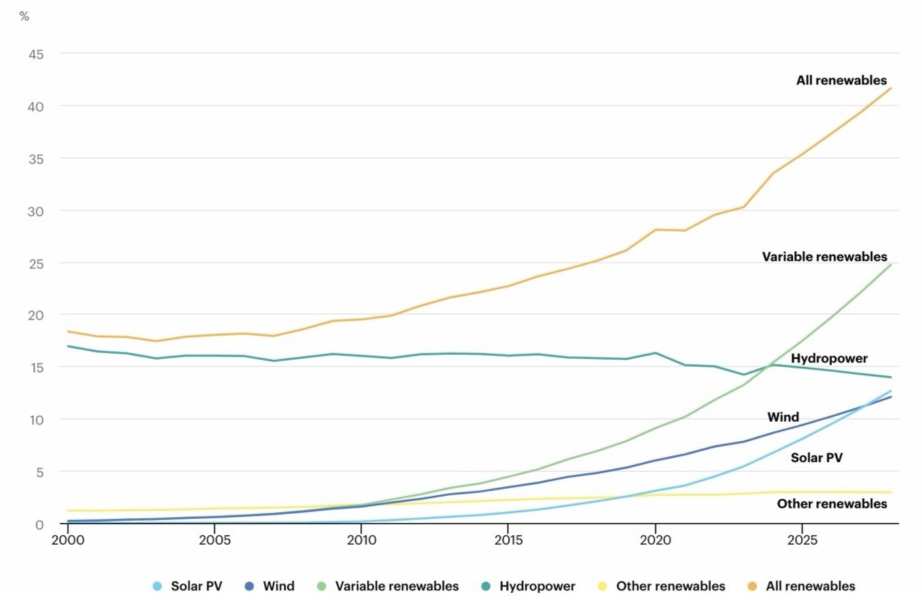

Figure 4. Share of renewable electricity generation 2000-2028. (IEA, 2024a)

In the future, variable renewable energy sources, i.e. wind and solar PV will play very important in global energy production as shown in figure 4. Wind and solar PV is projected to increase significantly in the future, while the production of other renewable energy sources will remain at the same level or even decrease.

Renewable energy sources are very diverse, but what wind and solar PV have in common is that they are often very unstable and intermittent. For this reason, improving the efficiency of renewable energy production and distribution, as well as developing energy storage technologies, is particularly important. Previous study by Ullah et al. (2022) focused to the generation of the renewable electricity production utilizing long term storage of renewable energy using the latest P2X technology. In the study, they found that P2X technology can store renewable energy for longer periods of time by the generation of H2 from the electrolysis process and CH4 from the methanation process. This thesis focuses on how the efficiency of production and distribution in renewable energy can be improved by using IoT and Big Data.

Smart grid facilitates the use of renewable energy sources in energy distribution. The concept of smart grids has been emerged to improve the flexibility and efficiency of the traditional grid, and enable electricity generation and distribution based on wind, solar and other renewable energy sources. the SG is different from a traditional grid in that it enables two-way communication and interactions between the devices and participants in supply chain. The two-way communication between the generation and consumption side is enabled by cooperation between different smart devices, meters, and energy storage systems. (Ullah et al., 2021)

According to IEA (2024b) wind and solar energy are the dominant RES of power generation to reach the net zero emission by 2050. It is particularly important that these two renewable energy sources are invested in, and their capacity is increased to achieve the goal. The reduction in the economic costs of renewable energy is very important for increasing the use of renewable energy sources in energy production and for getting rid of fossil fuels. Renewable energy sources such as solar and wind energy enable distributed energy production, where energy is produced as close as possible to its place of consumption, which can improve energy efficiency and reduce transmission costs significantly.

4.1 Wind Energy

Wind energy is an excellent source of renewable energy due to its low carbon footprint, high availability, and high technological readiness level. According to IEA (2024b), in 2022, 2100 TWh of electricity were generated with wind energy. Wind energy uses the kinetic energy of moving air, i.e. wind, and converts it into mechanical energy through the rotation of turbine blades, and then into electricity through generators. In this thesis we focus on three-bladed horizontal-axis wind turbines, both on- and offshore. According to Ren et al. (2021), offshore wind turbines have become the focus of wind technology development because they have many advantages compared to onshore wind turbines, such as, abundant wind resources, lower turbulence, substantial space for establishment, lower transmission and distribution losses, less visual impact, and less noise pollution. However, the downside of offshore wind turbines is significantly higher electricity production costs. The cost of electricity production with offshore wind turbines could be reduced by installing larger turbines further from the coast, but on the other hand, this would cause problems with the durability of wind turbines and thus would cause increase in maintenance and repair costs. (Ren et al., 2021)

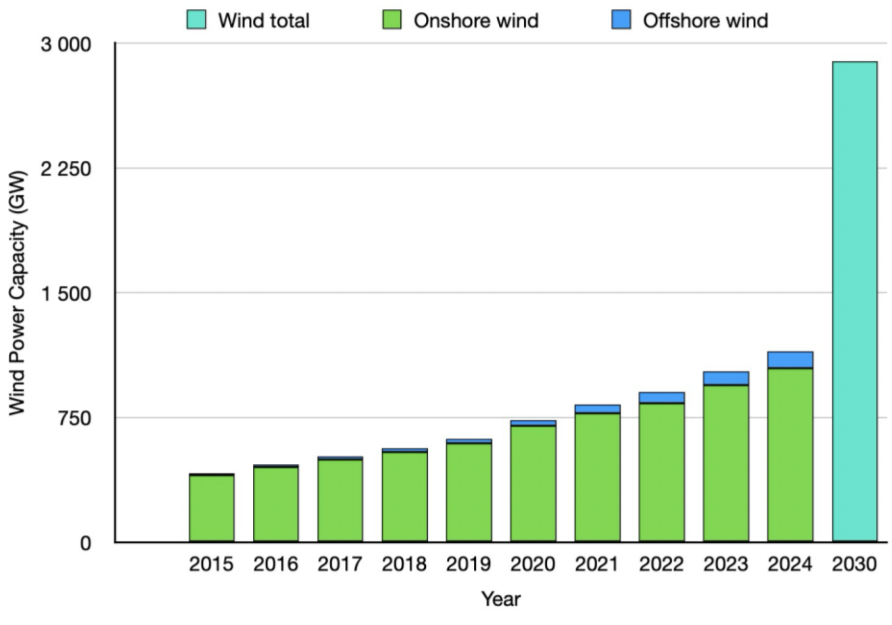

Figure 5. Wind power capacity in GW (IEA, 2024b)

As we can see from figure 5, wind power capacity still needs to be increased significantly in order to reach the 45 % of the emissions reduced target by 2030. To increase capacity, more political support is needed for wind power, with a particular focus on offshore wind power. Offshore wind power has a lot of potential, but due to its high LCOE, construction has fallen short. In order to increase capacity, the costs of offshore wind power must be reduced for example by applying IoT sensor-based predictive maintenance.

Operating and maintenance costs can be 23% of the total investment cost in an offshore wind turbine and 5% in an onshore wind turbine. Especially in an offshore wind turbine which are exposed to harsh conditions, maintenance costs play a major role. High maintenance costs are a big constraint especially for the development of offshore wind power. Section 6 of this thesis explores in more detail how IoT and Big Data can be used for predictive and condition-based maintenance to reduce maintenance costs in wind power. (Ren et al., 2021)

4.2 Solar Energy

Solar energy is based on converting sunlight into electrical energy by using either photovoltaics (PV) which is direct conversion of sunlight into electrical energy or CSP technologies which use mirrors to concentrate sunlight to receiver (UN, 2024). In this thesis we are focusing to the solar PV. According to IEA (2024), solar PV has grown tremendously in popularity over the past few years, while increasing its economic attractiveness. In 2022 solar PV generation have increased 270 TWh from 2021, reaching almost 1300 TWh. High electricity prices have increased the amount of decentralized solar PV for example on the rooftops of consumers' houses (IEA, 2024c). Like wind energy, the challenge with solar energy is its intermittent nature, as well as the fact that in some countries direct sunlight can be very limited during the day at the certain times of the year.

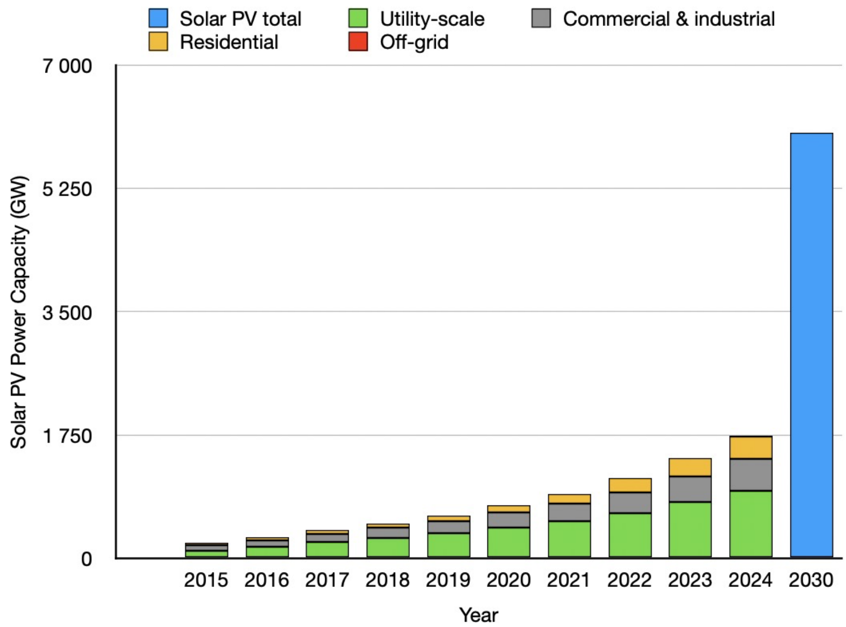

Figure 6. Solar PV power capacity in GW (IEA, 2024c)

Figure 6 represents solar PV power capacity in gigawatts, where Solar PV total is the capacity needed to reach the 45 % of the emissions reduced target by 2030. According to IEA (2024c), especially China, USA, EU and India, are expected to accelerate the solar PV capacity growth by increasing the policy support and developing the supply chain, which increases the economic attractiveness of solar PV. Utility-scale solar PV covers half of the total capacity, but distributed solar PV have been on the rise due to high fuel and electricity prices in past years. To achieve 45 % of the emissions reduced target by 2030, policy support and investments are needed for all solar PV segments. Solar PV production can be made more efficient and low-cost by utilizing IoT and Big Data to increase economic attractiveness of PV.

5 Working relationship between IoT and Big Data

The purpose of the Internet of Things is to connect different devices to each other and enable the communication between them and the platform. Various IoT devices and sensors are constantly collecting large amounts of heterogenous data. Big Data analytics can be used to process and store large amounts of complex and unstructured data and perform fast analytics to provide rapid insights and to facilitate the decision-making process. Implementing Big Data in IoT applications is extremely important, and these two technologies have already been recognized in the fields of IT and business. (Marjani et al., 2017)

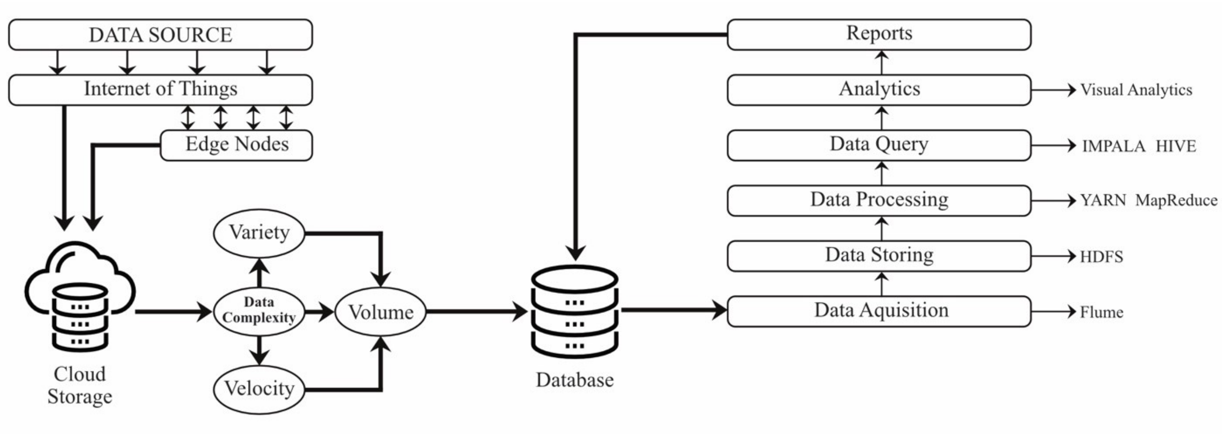

Figure 7. Big Data combined with IoT application. Adapted from (Ullah et al., 2021)

Figure 7 represents an IoT application utilizing edge computing and Big Data analytics. In their study, Ullah et al. (2021) have listed five main requirements of Big Data and analytics in IoT, which are: connectivity, storage, quality of services, real time analytics and benchmark. The main requirements are explained below.

1. *Connectivity:* With so many different devices on the network, connectivity is especially important. Reliable Big Data and IoT connectivity enables efficient integration of large masses of machine generated sensor data. Many devices

around us can connect to the computing and high-performance infrastructure through wireless networks and facilitate the IoT services. (Ullah et al., 2021)

2. *Storage:* As the amount of data produced continues to grow enormously, there is a need for a low-cost data storage. Big Data storage in IoT needs to be able to deal with huge amounts of structured, unstructured, and semi-structured data and provide low-latency analytics. (Ullah et al., 2021)
3. *Quality of Services:* Quality of Services (QoS) means an ability to provide a specific level of performance to the data flow. To meet the requirements of QoS, the IoT network must be reliable and provide an efficient data transfer from the data sources. (Ullah et al., 2021)
4. *Real time analytics:* Real time analytics is very important because in some IoT applications real-time communication and analyzation of data can be crucial. Big Data uses an operational database for streaming data. Big Data analytics performs real time queries for most of the streaming data to achieve fast information extraction for decision making. (Ullah et al., 2021)
5. *Benchmark:* The large amount of data generated by the IoT devices can cause difficulties in storing and analysing data for organizations that do not have a deep understanding of the problems. Benchmarking can be used to compare Big Data and analytics solutions. (Ullah et al., 2021)

Fulfilling these requirements enables the efficient combination and use of IoT and Big Data in variety of applications. IoT and Big Data combined provide the ability to store and analyse huge amounts of data gathered by the IoT devices very quickly to extract meaningful information from the data and improve the decision-making process (Ullah et al., 2021).

6 IoT and Big Data in Renewable Energy

IoT and Big Data play an important role in the field of renewable energy. IoT and Big Data enable the two-way interaction of smart grid that enables efficient usage of renewable energy in electricity production and distribution. In addition, IoT and Big Data can be used in many ways to enhance the production and distribution of renewable energy, such as wind and solar energy, while reducing costs at the same time. The purpose of this section is to compile the main uses of IoT and Big Data in renewable energy and justify how they improve efficiency and/or reduce costs.

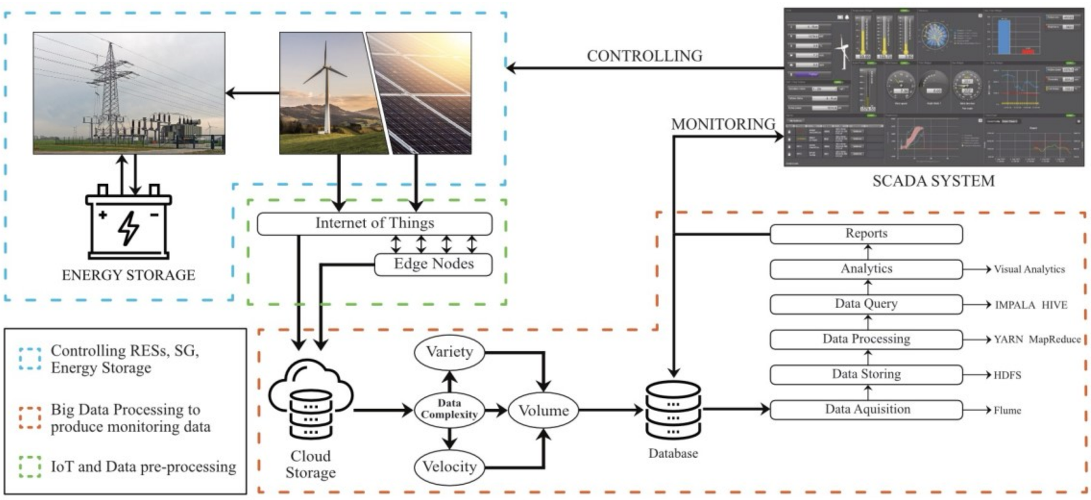

Figure 8. IoT and Big Data in Renewable Energy monitoring and controlling system by using SCADA.

The architecture how IoT and Big Data are involved in enabling the remote monitoring and controlling of renewable energy sources using SCADA system is shown in figure 8. The process starts from renewable energy sources where IoT sensors continuously gather data from RESs. The gathered data is preprocessed at the edge nodes and sent to the cloud storage using selected communication technologies. The huge amount of complex heterogenous data or Big Data is then transferred into shared distributed fault tolerant database, where the Big Data gets processed and analyzed as described in section 3.1. The SCADA system has access to the reports and processed data in the database which provides information that enables it to monitor and control renewable energy sources.

6.1 Monitoring and Controlling

Renewable energy production is highly distributed, and its production plants are often located in remote areas. Especially wind energy plants, as the turbines can be visually and acoustically disturbing to some people. IoT and Big Data enable remote monitoring and controlling of renewable energy production plants without human interventions, for example, with SCADA system. IoT based system SCADA, allows organizations to access the data generated by sensors in renewable energy sources and manage the production of solar PV and wind energy (SCADA-International, 2024).

In renewable energy production, the SCADA system and various IoT sensors can be used to measure various parameters of wind turbines, solar panels, and their environment. Parameters measured in wind energy production can be for example: wind speed and direction, voltage, current, power, humidity, and vibration. In solar power, on the other hand: weather conditions, voltage, current, panel temperature, and operating time. In renewable energy production, reliability is a very important factor in ensuring production efficiency. Remotely readable data from renewable energy sources can be used to identify faults, inefficiencies, and the need of maintenance at low cost. Remote monitoring and control can also reduce operating costs significantly. IoT can be used to automate the energy production process, for example, through the automatic rotation of solar panels or wind turbines to maintain maximum production efficiency. All applications of IoT and Big Data in renewable energy are based on this data collection and analysing process, which can be used to obtain valuable information of the production and distribution of renewable energy to make better and possibly automated decisions. (Ponnalagarsamy, 2021)

6.2 Production Optimization and Forecasting

One of the biggest challenges in renewable energy production is the intermittent nature of renewable energy sources, which can lead to a mismatch between supply and demand. IoT and Big Data technologies offer ways to optimize production and distribution by balancing renewable energy production and demand using forecasting based on machine learning (ML) technology.

Accurate forecasting of energy production and demand is mandatory to enable efficient energy management. The forecasting of energy production and demand is based on a huge amount of real-time and historical data produced by IoT devices and sensors, which is analyzed with ML to make predictions. Forecasting wind energy, solar energy, and load by using ML has three steps, which are: data preprocessing, model training, and forecasting. In data preprocessing important data is filtered and divided into training, validation, and testing datasets. Model training is used to create the functional forecasting model which can be used in the final phase where forecast is executed, and the results are visualized. Forecasting can be performed on several different time scales, short, medium, and long-term, which are all needed for efficient and reliable energy control. (Aslam et al., 2021)

By forecasting renewable energy production and load of the grid, the generated renewable energy can be used efficiently to match the demand. During times when energy demand is lower, the generated energy can be transferred to an energy storage, where it can be used later when demand increases. Production optimization and forecasting makes the electricity grid more flexible and increases the efficiency of utilizing renewable energy.

6.3 Maintenance

Maintenance plays a particularly important role with renewable energy production plants located in remote areas, and especially in offshore wind power. IoT and Big Data in renewable energy enable condition-based and predictive maintenance for solar panels and wind turbines. Solar panels do not usually require much maintenance during their lifetime, which is why condition-based and predictive maintenance plays a more important role in wind power. According to McKinsey & Company (2015), utilizing Big Data analytics in IoT enabled machines can reduce maintenance costs up to 40 %. This is especially important in offshore wind power, where operating and maintenance (O&M) costs can be up to 23 % of the total investment cost, compared to onshore wind turbines where the O&M costs is around 5 % of the total investment costs (Ren et al., 2021).

Reducing O&M costs is an effective way to reduce levelized cost of energy (LCOE), which stands for the average life-cycle price of the electricity generated from a given power source in megawatt-hours. Offshore wind power has a relatively high LCOE, which has been attempted to be solved by building larger wind turbines farther out to the sea, which in the other hand increases failure rates, and repair and maintenance costs. Offshore wind turbines are exposed to harsher conditions, which can easily lead to increased maintenance costs. Maintenance has major impact on downtime, and therefore to the production efficiency, as well as to the LCOE. (Ren et al., 2021)

IoT sensors and Big Data analytics enable condition-based and predictive maintenance in renewable energy production. In condition-based maintenance, various sensors monitor the condition of components and try to identify when a component is worn out to prevent major failures from happening. Predictive maintenance is a bit more advanced and uses sensor measurements and data analyses to determine when the maintenance should be performed, so the maintenance can be done before any failure occurs. In wind turbines, the need for maintenance can be measured, for example, through vibration, oil, sound, and performance analyses. (Ren et al., 2021)

IoT and Big Data can be used to collect data from renewable energy production and extract valuable information about need of maintenance to increase efficiency and reduce costs of the energy production. By using sensor-based maintenance reliability of the production is increased, downtime is minimized, and maintenance costs are reduced.

7 Conclusions

This thesis explained Internet of Things, Big Data, and renewable energy concepts, and how they work together. The main objective of this thesis was to find out how IoT and Big Data can be utilised in improving the efficiency of renewable energy production and distribution. We also discussed how the use of these technologies can reduce costs of the renewable energy production and distribution. The study found that IoT and Big Data can be used to increase efficiency and reduce costs in renewable energy, as these technologies can enable remote monitoring and management of production plants, production and demand forecasting and sensor-based predictive maintenance. As we are living a revolution of artificial intelligence, we can assume that as AI develops, it will have increasingly important role in Big Data processing, and it will provide extremely accurate analyses based on the huge amount of data collected by the IoT devices.

Although IoT and Big Data should definitely be utilised in the field of renewable energy, they also have certain challenges and limitations which needs to be considered. Utilizing IoT increases the complexity of systems, increases investment costs, and if poorly implemented, can pose significant security and privacy risks. However, when implemented well, IoT utilization has significant benefits in improving efficiency and reducing costs in the long run, so the downsides should not be overweighted.

Further research could focus on, for example, more specifically to artificial intelligence and machine learning solutions, energy storage (P2X) technologies or examine IoT-based renewable energy applications more from a consumer perspective. In any case, IoT applications need to be researched further, as the rapid development of technology constantly brings new opportunities and challenges where to utilize IoT.

References

Apache. 2023. Apache Hadoop. [Online]. [Accessed 13 December 2023]. Available: https://hadoop.apache.org/

Apache .2024a. Apache Flume. [Online]. [Accessed 4 January 2024]. Available: https://flume.apache.org/

Apache. 2024b. Apache Hadoop YARN. [Online]. [Accessed 4 January 2024]. Available: https://hadoop.apache.org/docs/stable/hadoop-yarn/hadoop-yarn-site/YARN.html

Aslam S., Herodotos H., Mohsin S. M., Javaid N., Ashraf N., Aslam S. 2021. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renewable & sustainable energy reviews*. [Online]. 144110992-. [Accessed 17 January 2024]. Available: https://doi.org/10.1016/j.rser.2021.110992

Bifet A. 2015. SAMOA: "Scalable Advanced Massive Online Analysis." in *2015 Journal of Machine Learning Research*. [Online]. vol 16. pp. 149-153. [Accessed 4 January 2024]. Available: https://www.jmlr.org/papers/volume16/morales15a/morales15a.pdf

Çam E., Alvarez C., 2023. Europe's energy crisis: Understanding the drivers of the fall in electricity demand. Commentary. IEA. [Online]. [Accessed 14 November 2023]. Available: https://www.iea.org/commentaries/europe-s-energy-crisis-understanding-the-drivers-of-the-fall-in-electricity-demand

Dong F., Yong J., Fei X. 2020. Emerging intelligent big data analytics for cloud and edge computing. *Concurrency and computation*. [Online]. 32 (23). [Accessed 30 December 2023]. Available: https://doi.org/10.1002/cpe.5989

Environmental and Energy Study Institute. 2021. Fact sheet: Climate, Environmental, and Health Impacts of Fossil Fuels. [Online]. [Accessed 12 November 2023]. Available: https://www.eesi.org/papers/view/fact-sheet-climate-environmental-and-health-impacts-of-fossil-fuels-2021.

European Wind Energy Association. 2024. Wind energy frequently asked question. [Online]. [Accessed 6 January 2024]. Available: https://www.ewea.org/wind-energy-basics/faq/

Google Cloud. 2023. What is Big Data? Google Cloud learn. [Online]. [Accessed 13 December 2023]. Available: https://cloud.google.com/learn/what-is-big-data

Greer C., Burns M., Wollman D., Griffor E. 2019. Cyber-Physical Systems and Internet of Things. *NIST Special Publication 1900-202*. NIST, USA, 2019. [Online]. [Accessed 20 November 2023]. Available: https://doi.org/10.6028/NIST.SP.1900-202

Hejazi H., Rajab H., Cinkler T., Lengyel L. 2018. "Survey of platforms for massive IoT." in *2018 IEEE International Conference on Future IoT Technologies (Future IoT).* [Online]. 2018 IEEE. pp. 1–8. [Accessed 22 November 2023]. Available: https://doi.org/10.1109/FIOT.2018.8325598

International Energy Agency. 2023a. World Energy Outlook 2023. IEA, Paris 2023. [Online]. [Accessed 10 November 2023]. Available: https://www.iea.org/reports/world-energy-outlook-2023

International Energy Agency. 2023b. Global Energy Crisis. [Online]. [Accessed 13 November 2023]. Available: https://www.iea.org/topics/global-energy-crisis

International Energy Agency. 2024a. Renewables. [Online]. [Accessed 5 January 2024]. Available: https://www.iea.org/energy-system/renewables

International Energy Agency. 2024b. Wind. [Online]. [Accessed 6 January 2024]. Available: https://www.iea.org/energy-system/renewables/wind

International Energy Agency. 2024c. Solar PV. [Online]. [Accessed 9 January 2024]. Available: https://www.iea.org/energy-system/renewables/solar-pv

Jadeja, Y. & Modi, K. 2012. "Cloud computing - concepts, architecture and challenges." in *2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET)*. [Online]. 2012 IEEE. pp. 877–880. [Accessed 30 December 2023]. Available: https://doi.org/10.1109/ICCEET.2012.6203873

Lee I. & Lee K. 2015. The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business horizons.* [Online]. 58 (4), 431-440. [Accessed 12 December 2023]. Available: https://doi.org/10.1016/j.bushor.2015.03.008

Mandal S. 2020. Internet of Things (IoT) – Part 1 (Introduction). [Online]. [Accessed 16 November 2023]. Available: https://www.c-sharpcorner.com/UploadFile/f88748/internet-of-things-iot-an-introduction/

Marjani M., Nasaruddin F., Gani A., Karim A., Hashem I. A. T., Siddiqa A., Yaqoob I. 2017. Big IoT Data Analytics: Achitecture, Opportunities, and Open Research Challenges. *IEEE Access*. [Online]. 55247-5261. [Accessed 4 January 2024]. Available: https://doi.org/10.1109/ACCESS.2017.2689040

McKinsey Global Institute. 2015. The Internet of Things: Mapping the Value Beyond the Hype. [Online]. [Accessed 5 January 2024]. Available: https://www.mckinsey.com/~/media/The-Internet-of-things-Mapping-the-value-beyond-the-hype.pdf

Microsoft. 2023. What is edge computing? [Online]. [Accessed 1 January 2024]. Available: https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-edge-computing

Ponnalagarsamy S., Geetha V., Pushpavalli M., Abirami P. 2021. Impact of IoT on Renewable Energy. *IoT Applications Computing*. [Online]. pp. 107-117. [Accessed 12 December 2023]. Available: https://doi.org/10.5772/intechopen.98320

Raguseo E. 2018. Big Data technologies: An empirical investigation on their adoption, benefits and risks for companies. *International journal of information management*. [Online]. 38 (1), 187-195. [Accessed 5 January 2024]. Available: https://doi.org/10.1016/j.ijinfomgt.2017.07.008

Ren Z., Verma A. S., Li Y., Teuwen J. J. E., Jiang Z. 2021. Offshore wind turbine operations and maintenance: A state-of-the-art review. *Renewable & sustainable energy reviews.* [Online]. vol. 114. 110886-. [Accessed 6 January 2024]. Available: https://doi.org/10.1016/j.rser.2021.110886

Rhayem A., Mhiri M. B. A., Gargouri F. 2020. Semantic Web Technologies for the Internet of Things: Systematic Literature Review. *Internet of Things.* [Online]. vol. 11. 11100206- [Accessed 22 November 2023]. Available: https://doi.org/10.1016/j.iot.2020.100206

SAP. 2023. What is Big Data? [Online]. [Accessed 4 January 2024]. Available: https://www.sap.com/uk/products/technology-platform/what-is-big-data.html

SCADA-International. 2024. What is SCADA? [Online]. [Accessed 15 January 2024]. Available: https://scada-international.com/what-is-scada/

Sehrawat D. & Gill N. S. 2019. "Smart Sensors: Analysis of Different Types of IoT Sensors." in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI).* [Online]. 2019 IEEE. pp. 523-528. [Accessed 22 November 2023]. Available: https://doi.org/10.1109/ICOEI.2019.8862778

Shah S. H. & Yaqoob I. 2016. "A Survey: Internet of Things (IOT) Technologies, Applications and Challenges." in *2016 IEEE Smart Energy Grid Engineering (SEGE)*. [Online]. 2016 IEEE. pp. 381-385 [Accessed 24 November 2023]. Available: https://doi.org/10.1109/SEGE.2016.7589556

Shi W., Cao J., Zhang Q., Li Y., Xu L. 2016. Edge Computing: Vision and Challenges. *IEEE internet of things journal*. [Online]. 3 (5), 637-646. [Accessed 25 November 2023]. Available: https://doi.org/10.1109/JIOT.2016.2579198

Tableau. 2023. Big Data Analytics: What It Is, How It Works, Benefits, And Challenges. [Online]. [Accessed 30 December 2023]. Available: https://www.tableau.com/learn/articles/big-data-analytics

Tyagi A. K., Abraham A., Hussain F. K., Kaklauskas A., Kannan R. J. 2022. Machine Learning, Blockchain Technologies and Big Data Analytics for IoTs – Methods, Technologies, and Applications. *Stevenage: Institution of Engineering & Technology*. First edition. pp. 75-85. [Accessed 14 December 2023].

Ullah M., Gutierrez-Rojas D., Inkeri E., Tynjälä T., Nardelli P. H. J. 2022. Operation of Power-to-X cogeneration plants based on advanced data-driven methods: A comprehensive review. *Energies* 2022. [Online]. 15 (21), 8118. [Accessed 24 November 2023]. Available: https://doi.org/10.3390/en15218118

Ullah M., Narayanan A., Wolff A., Nardelli P.H.J. 2021. "Smart Grid Information Processes Using IoT and Big Data with Cloud and Edge Computing." In *2021 44th International Convention on Information, Communication and Electronic Technology*

(MIPRO). [Online]. 2021 Croatian Society MIPRO. pp. 956-961. [Accessed 20 November 2023]. Available: https://doi.org/10.23919/MIPRO52101.2021.9596885

Ullah M., Nardelli P. H. J., Wolff A. Smolander K. 2020. Twenty-one key factors to choose an IoT platform: Theoretical framework and its applications. [Online]. [Accessed 14 November 2023]. Available: https://doi.org/10.1109/JIOT.2020.3000056

United Nations. 2024. For a livable climate: Net-zero commitments must be backed by credible action. [Online]. [Accessed 5 January 2024]. Available: https://www.un.org/en/climatechange/net-zero-coalition

Yaqoob I., Hashem I. A. T., Gani A., Mokhtar S., Ahmed E., Anuar N. B., Vasilakos A. V. 2016. Big Data: From beginning to future. *International journal of information management.* [Online]. 36 (6), 1231-1247. [Accessed 13 December 2023]. Available: https://doi.org/10.1016/j.ijinfomgt.2016.07.009