Big Data and IoT in smart homes

Patricia Mezani[a],Nirida Pashaj[,b]and Romina Agaçi[c]

[a] *University of Vlora "Ismail Qemali", Skele, Rruga Kosova, Vlore, 9401, Albania*
[b] *University of Vlora "IsmailQemali", Skele, Rruga Kosova, Vlore, 9401, Albania*
[c] *University of Vlora "IsmailQemali", Skele, Rruga Kosova, Vlore, 9401, Albania*

Abstract

The term "big data" refers data that are big, fast or complex that it is difficult or impossible to process using traditional methods. Thanks to the invention of mobile technology like smartphones and tablets, along with innovations in mobile networks and Wi-Fi, the creation and consumption of data are constantly growing. It is not necessary the amount of data you have, but what you can do with data. You collect data analyze them for a better cost and time reduction, to develop new products and what is more important to make smart decision. "Data" is termed as the new oil in the 21st century. IoT is about devices, data and connectivity. Internet of Things (IoT) is software that connects the edge hardware, access points, and data networks to the end-user application. Millions of devices are connected together, so internet of things will trigger a massive inflow of Big Data. In this paper, we briefly review the connection of Big Data and IoT in an open source platform like Open Remote that integrate all your data, sensors and controls to access them centrally, visualizing and analyzing your data to gain new insights and always stay in control.

1. Introduction

The size and number of available data sets has grown rapidly .As it said[1] data never sleeps

In 2020, the world changed fundamentally .As Covid-19 swept the globe, nearly every aspect of life from work to working out moved online, and people depended more and more on apps and the Internet to socialize, educate and entertain ourselves. Consumer Internet of Things applications can range from very simple and cheap ones such as personal fitness devices to high end smart home automation applications.[2]

The aim of this paper is to provide an overview of what is big data, IoT, the connection between them and opportunities, challenges and technologies that are using today especially in smart homes. The main contributions of this paper can be summarized as follows

- We provide a summary on key characteristics of Big Data and IoT.
- We discuss necessities and challenges of big data.
- We present an overview of technologies used for big Data and IoT
- We present a practical view in smart devices of smart homes.

2. Characteristics of Big Data

There are three types of Big Data; structured, unstructured and semi-structured. Any data that can be stored, accessed and processed in the form of fixed format is termed as structured data. Data stored in a relational database management is one example of a structured data. Any data with unknown form or the structure is classified as unstructured data. A typical example of unstructured data is a data source that contain text files, images, videos, for example the output returned by google search. While semi-structured data can contain both the forms of data for example a table definition in relational DBMS[3].There are different classification of characteristics of Big Data . The 4V's of Big data[4] include

- Volume: The quantity of generated and stored data. The size of data determines the value and potential insight and whether it can be considered big or not

- Variety: The type and nature of the data
- Velocity: The speed at which the data is generated and processed. Big data is often available in real time. Compared to small data, big data are produced more continually
- Veracity: Data quality and data value.
  Nowadays are using others V's for definition of Big Data like as value, variability etc.

2.1 What is IoT ?

The smart devices you see every day are a small part of IoT. Your smart television is an IoT device. You can talk to it using an app on your smartphone. You tell it to turn on and you can see your favorite program. As is shown in Figure 1 we have big data everywhere. A smart home[5] has appliances lighting, heating, air conditioning, TVs, computers, entertainment systems, big home appliances, washers/dryers and refrigerators/freezers, security and camera systems capable of communicating with each other and being controlled remotely by a time schedule, phone, mobile or internet. IoT devices can be used to monitor and control the mechanical, electrical and electronic systems used in various types of buildings.

2.2 The importance of Big Data

Big Data use different techniques like machine learning, data mining, natural language processing and statistics. But why Big Data is so important and efficient? There are three main reasons why Big Data is so important[6].

1. **Cost Reduction**. Big Data technologies such as Hadoop and cloud based analytics bring significant cost advantages when it comes to storing large amounts of data.
2. **Faster, better decision making**. With the speed of Hadoop and in-memory analytics, combined with the ability to analyze new sources of data, business is able to analyze information immediately and make decisions based on what they have learned
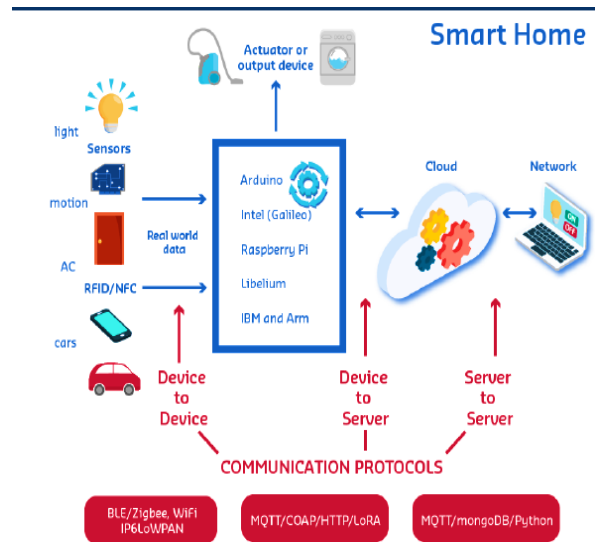

Figure 1 Big Data Everywhere

3. **New products and services**. Consumers take what they want[7]

2.3 Challenges of Big Data

Often has insufficient understanding and acceptance of big data. Big data adoption projects entail lots of expenses. You have costs for devices, hiring, electricity and so on. Another challenge is complexity of managing data quality. You have data from diverse sources and unreliable data. Also you sooner or
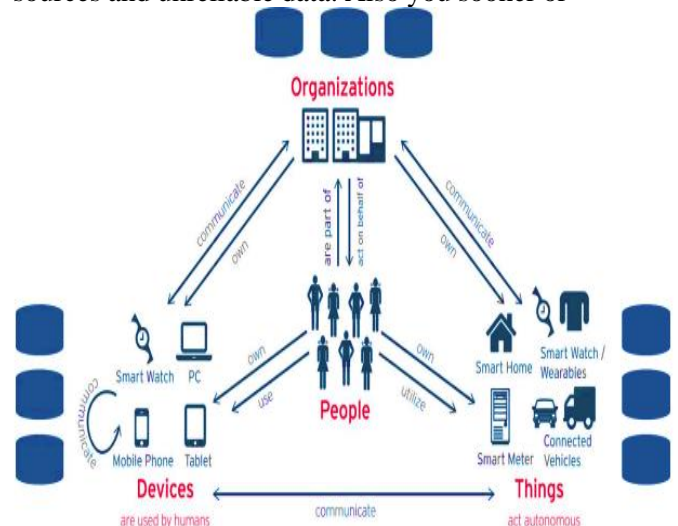
later, will run into the problem of data integration, since the data you need to analyze comes from diverse sources in a variety of different formats. For instance, ecommerce companies need to analyze data from website logs, call-center, competitors website scans and social media. The Big Data isn't 100% accurate. Security challenges of big data is also

another challenge. The most typical feature of big data is its dramatic ability to grow. And one of the most serious challenges of big data is associated exactly with this.

3. Big Data Statistics

3.1 Big Data 2019 Statistics

Although the term big data has been known since the 1960s, in recent years it is taking more meaning. Since 2005, the year when social media started to grow and have more and more users, therefore more and more information was circulating, this term would take on a different value. In fact, as "data never sleeps" can often be read, it would be the expression that best defines and describes what happens to data today. We create 2.5 quintillion bytes[8]of data every day.

According to DOMO -in 2019, some platforms that circulate more data per minute are as follows[9]

Table 1
Data Flow : 2019

Platform	The amount of data per minute	Unit
Google	4 497 420	Searches
Email	188 000 000	Sent
YouTube	4 500 000	Videos watched
Giphy	4 800 000	Serves up gifs
Instagram	333 917	Posts/photos
Text	18 100 000	Are sent
Tinder User swipe	1 400 000	times
Netflix	694 444	video

3.2 Big Data 2020 Statistics

In 2020 google remains the platform that circulates the most daily data .Google gets over 3.5 billion searches daily.[10]

And what is worth noting is that 15% of searches are new searches, searches that have never been typed before.

Which means that is not generated the same information.

For 2020Whatsapp was the most downloaded messaging application and generated the most data. 65 billion messages [11]are sent daily through this app

YouTube was launched in 2005. Since the upload of the first video "Me at the zoo", this platform has experienced a very fast growth. Today, 500 hours [12]of content are uploaded to the platform every minute.

Email is another platform that circulates a large amount of data. In 2020, 306.4 billion emails are sent daily

Zoom is one of the platforms that got great growth in 2020. Thus in 2020 in zoom 300 million meetings were made per day.

[13]Briefly in 2018, the amount of data created every day was almost 2.5 quintillion bytes.

In 2020 the amount of data in the world was estimated to be 44 zettabytes.

3.3 The cost of poor data quality

Before Big Data analytics became a fully developed idea, companies were storing tons of info in their databases, not knowing what to do with them. Poor data quality causes poor or wrong decision making and wrong business strategies. They also directly affect productivity. Bad data is both time consuming and energy consuming.

Poor data quality is not just a problem for the company's revenue, it directly affects the company's image. According to global statistics on Big Data technologies, poor data quality costs businesses worldwide anywhere between $9.7 million and $14.2 million yearly.

4. The future of big data

Clearly with the increase of data a decade ago and until today the data are getting bigger. Obviously this growth is not going to stop Big data have grown and continue to grow through internet searches, texting, social networking, etc.

Another growth has occurred and will continue even more in the future, through IOT devices. It is already a fact that data will grow more and more. This big data will migrate to the Cloud. So by 2025, the amount of data generated each

day is expected to reach 463 Exabyte's globally.

By 2030, 90% of the population older than six will be online.

In summary, we can say that the future of big data will change the way businesses are organized, their strategies, and even the way of doing business.

However we must be prepared that such changes will definitely be associated with problems, one of the most important will be security and data privacy.

Despite all the challenges, uncertainties about the future, risks and potential problems, we can say that big data is big value.

5. Big data and Internet of Things platforms

Big data platform generally consists of big data storage, servers, database, big data management and business intelligence. Big data platform is a type of IT solution that combines the features and capabilities of several big data application and utilities within a single solution[14]. The IoT technology consists of multiple layers, including device hardware, connectivity, data management, applications, analytics and security[15]. The IoT platforms are suites of components that are used to setup, manage and automate the internet-connected devices. There are four types of IoT platforms: End to end, Connectivity, Cloud and Data. The role of big data in IoT is to process a large amount of data on a real-time basis and storing them using different storage technologies. Big data analytics tools have the capacity to handle large volumes of data generated from IoT devices that create a continuous stream of information. However, IoT provides data from which big data analytics can extract information to generate insights required of it. Hence, without the data, IoT devices would not hold the functionalities and capabilities, which have made them, achieve so much worldwide attention.

6. Home Automation

Home automation covers a variety of technological capabilities you can install in your home. It is a method of controlling home appliances automatically for the convenience of users. Home automation can include controlling aspects of your home remotely through a computer or phone, programming electronic devices to respond automatically to certain conditions or scenarios, or centralizing the control of a variety of items in your home into a single control center[16].This technology makes life easier for the user, and saves energy by utilizing devices according to strict requirements. Devices can be controlled by the user to suit personal requirements using cell phones, direct buttons, the internet, or infrared remotes. A network of appliances and sensors can interact with each other and make decisions for operation. If a user wants to control the home appliances through the Internet, the designer needs to add an Ethernet/Wi-Fi interface to connect the system to the home network. If the user wants to control the system using Bluetooth from a cell phone, the designer needs to add a Bluetooth interface to communicate with the device. The range of sensors that should be considered include temperature sensors, water sensors, LUX sensors, humidity sensors, pressure sensors, air composition sensors, video cameras for surveillance, voice/sound sensors, infrared sensors, accelerometers, vibrations sensors, and ultrasonic sensors. It is important to make the entire system secure so that it cannot be easily altered to give control of the house to unauthorized users.

6.1 OpenRemote in home automation

Open Remote[17] is a 100% open-source IoT platform to create a wide range of applications. It integrates all your data, sensors, and controls to access them centrally. Open Remote designs any application for your unique project and create a smooth workflow. It visualizes and analyses your data to gain new insights and always stay in control.

Open Remote enables you to connect virtually all types of devices within your network and create an intuitive universal remote on your tablet or smartphone, or do without a UI and automate everything using Rules. Through OpenRemote you can manage your scenes comfortably in your living room, combine lighting ambiance and climate.

With a simple interface, the user can create his/her own rules. These rules can describe

thresholds that, if crossed, trigger an alert. These alerts can be automatically directed to the one responsible for that asset by e-mail or in-app notification so that they can solve the problem before it occurs.

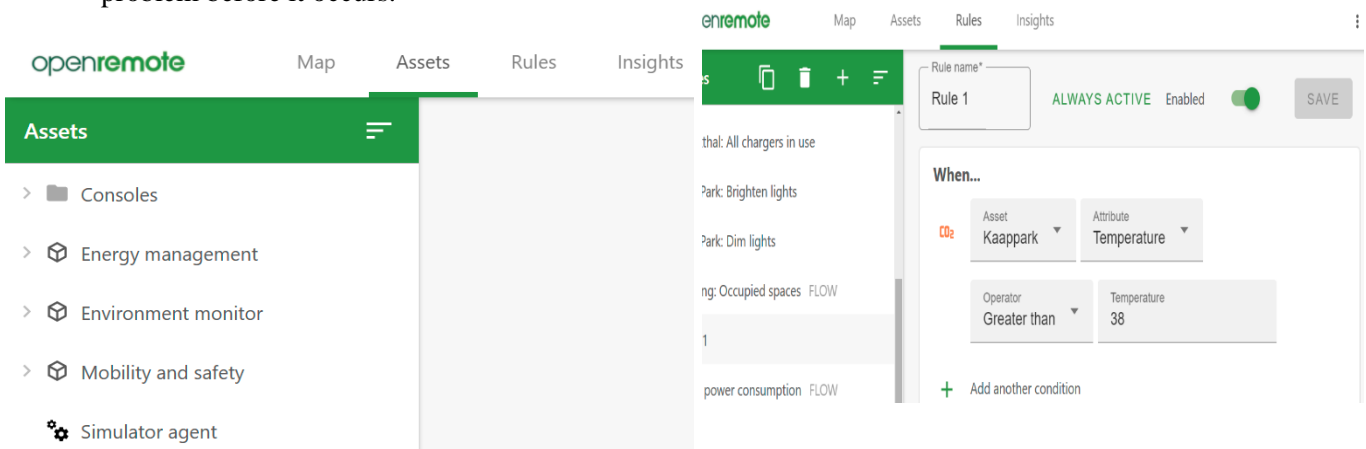With an easy user interface, you create rules to describe behavior of your system ("when-then" rules).
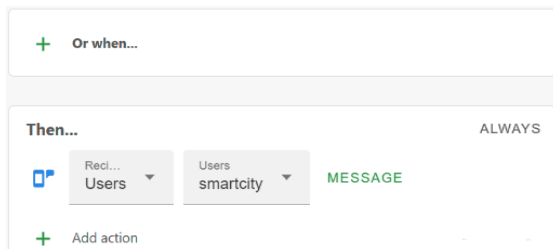

Figure 3 OpenRempte Assets

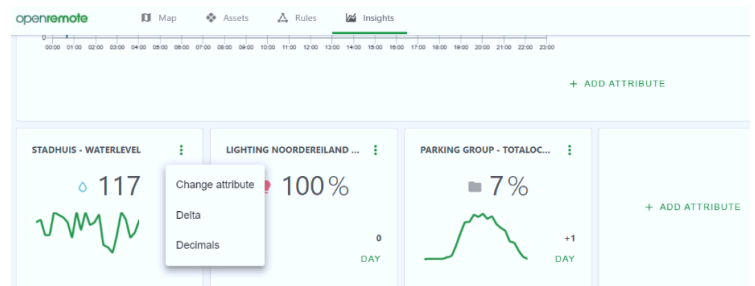

Figure 4 Open Remote Rules ("when-then")

Figure 5 Open Remote Insights

With all live values in one place, performance of assets can be reviewed on the Insights page.The freedom to create these graphs on the fly withwith live data lets you analyze the system very effectively

[1]https://www.domo.com/learn/data-never-sleeps-8

[2]https://www.i-scoop.eu/internet-of-things-guide/what-
is-consumer-internet-of-things-ciot/

[3]https://www.guru99.com/what-is-big-data.html

[4]D. Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety," Gartner Inc., 2001.

[5]https://www.pcmag.com/news/the-best-smart-home-devices

[6] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," Computer Networks, vol. 54, no. 15, pp. 2787–2805, 2010

REFERENCES:

[7]https://medium.com/@syedjunaid.h47/what-is-big-data-why-is-big-data-important-in-todays-era-8dbc9314fb0a

[8] Allan Jey,97 Big Data Statistics You Must Learn ,article for Finances online.

[9]Seventh edition of Data Never Sleeps, for DEMO ,2019

[10]Internet Live Stats

[11]Connective systems

[12]YouTube revenue and usage Statistics (2020), from Mansoor Iqbal , updated March 8,2021

[13]Seed scientific, "How much data is created every day"

[14]Using big data platforms for data management, access and analytics (techtarget.com)

[15] Proceedings of the 6th International Conference on Information Society and Technology (ICIST 2016) At: Kopaonik, Serbia. Survey of Internet-of-Things platforms, p.1

[16] Proceedings of the International Journal of Computer Applications. Smart Home Automation: A Literature Review. p.1-2.

[17] OpenRemote | The 100% Open Source IoT Platform