

# Big Data and IoT-based Applications in Smart Environmental Fields: A Systematic Review

---

## Abstract

This paper reviews big data and Internet of Things (IoT)-based applications in fields related to smart environments. This research is intended to identify the areas of application, current trends, data architectures, and ongoing challenges in this field. This article, which is the first systematic review of its kind, aims to achieve this goal by using draws on academic documents published in peer-reviewed venues from 2011 to 2019, based on a four-step selection process of identification, screening, eligibility, and inclusion for the selection process. In order to examine these articles, a systematic review was conducted and six main research questions were answered. The results indicate that the integration of big data and IoT technologies creates great opportunities for smart environment applications regarding monitoring, the protection, the monitoring, and the improvement of natural resources. The tools that have been developed in these areas include smart environment monitoring, smart farming/agriculture, smart metering, and smart disaster alert). We conclude by summarizing the methods most commonly used in big data and IoT, which we intend to serve as a starting point for future research.

*Keywords:* Systematic review; Internet of Things (IoT); big data; smart environment applications

---

## 1. Introduction

The internet of things (IoT) has risen in use and popularity over the last decade, pointing toward productive and exciting new directions for a whole generation of information devices [1] [2]. The fundamental concepts of IoT were invented by Kevin Ashton in 1999, when he introduced device-to-device communication on a wider scale than previously possible [3]. Atzori et al [4] have since stated that "IoT is a result of the convergence of three

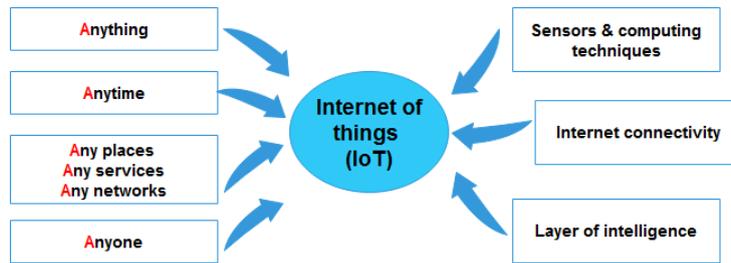


Figure 1: Internet of things Concept

visions; things-oriented, internet-oriented, and semantic-oriented”. In terms of semantics specifically, the IoT is a ”world-wide network of interconnected objects”. According to [5] [6], IoT can be defined as ”a dynamic global network infrastructure, as such it can identify, control, and monitor each object on earth via the internet according to a specific agreement protocol, and through the interconnection of physical and virtual things based on the interoperability of information and communication technologies.” The main purpose of IoT is to help share real-time information through autonomous networked actors [7]. Figure 1, explains the concept of IoT [8]. A sensor with intelligent computational capacities is placed in a location containing an internet connection. This sensor will be able to communicate with anything, at any time, and from any place within the network. Data collection systems locate and transfer data through large numbers of such communication devices within the IoT infrastructure, facilitating the process of data collection. Several communication solutions, such as WIFI, ZigBee, Bluetooth, and GSM, interconnect devices using various access networks including Radio Frequency Identification (RFID), devices with wireless sensors, and any smart object connected to the internet over physical IP [9].

Tremendous volumes of data are generated by IoT. These results, often termed ”big data,” refer to ”large scale of data that demands new architectures and technologies for data management (capturing and processing) to enable the extraction of value for enhanced insight and decision making” [10]. Big data is characterized by various high-volume, high-velocity, high-variety, and high-veracity properties [11] [12]. By 2020, IoT is expected to connect 50 billion or more devices, due to the extensive influx of new smart objects and exponentially rising demand for their services [6].

Recently, IoT has been applied in smart environments which allow users to better understand and control their environment through a range of interconnected devices [14]. In smart environment applications, IoT is employed in order to build a comprehensive, multi-level, and fully covered ecological monitoring network which can be achieved using integration sensors at all levels leveraging IoT with spatial and temporal information, and constructing a massive platform with a data center and unified service support [15]. IoT technology and its integration with big data have been applied extensively across diverse fields such as smart cities, smart healthcare, intelligent warning systems, disaster management, etc [16]. Therefore, the construction and application of IoT and big data in environmental fields has

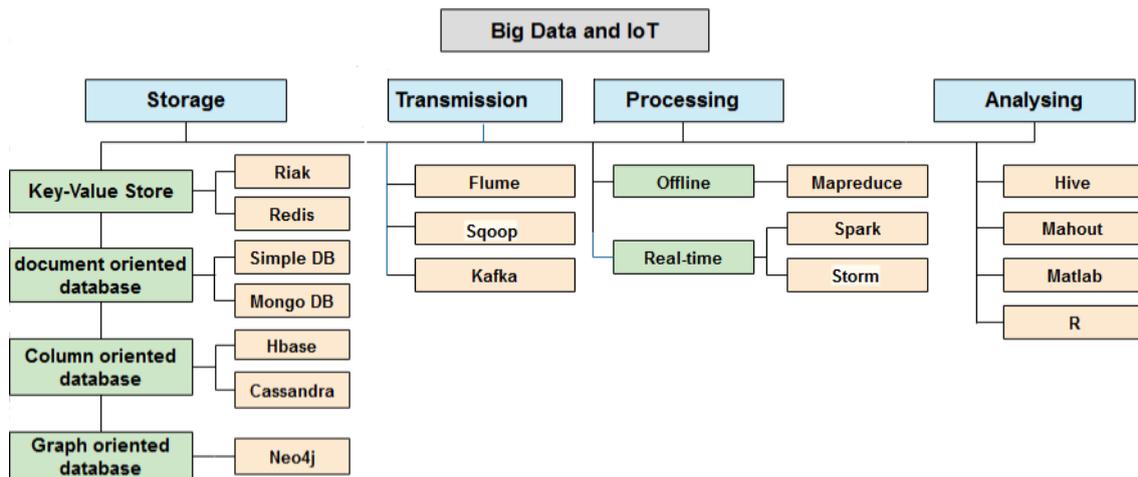


Figure 2: Main tools used in IoT and big data.

become a crucial measure, particularly for the development, promotion, and management of a strategic new environment protection industry [12].

Although significant effects have been achieved in IoT for smart-environmental applications, it can still be a challenging task to capture, store, process, and analyze the vast volumes of data computational intelligence techniques. To achieve the desired results, smart-environment applications require a real-time or even near real-time processing over high-speed data streams, which in turn requires fast, efficient, and large-scale data streaming analysis. The exponential rise in the volume and variety of new sensors available makes it particularly difficult for users to analyze, share, and extract knowledge from the big data. Big data and IoT-based applications in the field of smart environments is currently an under-examined area of research, and one that offers several interesting challenges and promising opportunities. The IoT offers a framework of ubiquitous networks with interlinked sensors, smart devices, and diverse data sources, while IoT technologies possess the potential for incorporation into existing applications as well as new insights for the development of future applications [17]. The expansion of communication made possible by Web 2.0 has been further driven by integration of different tools and applications. It is now possible to integrate, heterogeneous data-gathering sources (i.e., IoT enabled sensors, satellites, and social media) and powerful big data analytic tools (i.e. Hadoop, Spark, Kafka, etc.) with visualization applications (i.e. Kibana), and such possibilities have promising implications for the evolution of smart environment applications. Following our survey of the current research in big data and IoT, we have determined that big data technology could be used, for transmission, processing, storage and analysis purposes across the field of smart environment and its related application field. The main tools used in this area are shown in Figure 2.

### 1.1. Related Surveys

In this work, we survey the existing literature regarding big data and IoT within the scope of smart environment applications. Our main goal is to study the purpose of combining big data and IoT, which we undertake in order to explain how this combination can lead to promising solutions and new insights for smart environment applications, to explore the as-yet unrecognized opportunities and hidden challenges of such work, and to identify future research opportunities. The aim of this systematic review (SR) paper, is to assist fellow researchers and practitioners to better understand, and thus implement, the concepts of big data and IoT in this field.

In order to complete this research study, we have examined published surveys (e.g. systematic mapping study [SMS] [18], systematic literature review [SLR], and systematic reviews [SR]) related to big data and IoT and published in the field of smart environments from 2011 to 2019. The number of highest citation research articles with the most citations is shown in Figure 3. Then, a summary of the main objectives of each survey paper is given

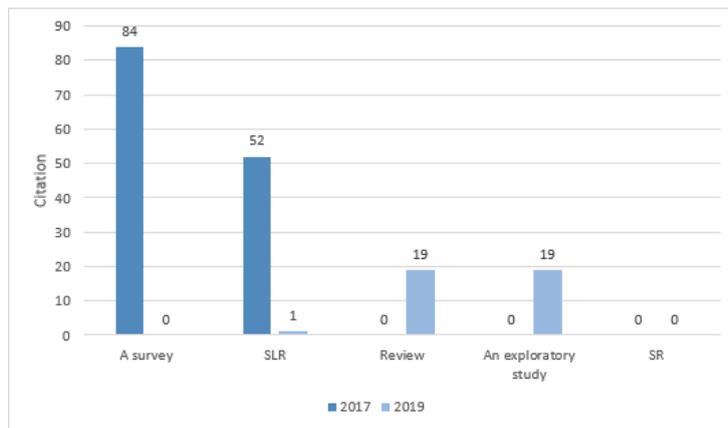


Figure 3: Distribution of most citation surveys articles per type

in Table 1. The list of main papers found verifies the search queries S3 and S4 (depicted in Table 4). Among 5 surveys, only 3 papers focus on the combination of big data and IoT in environment applications. Most recently, in 2019, Seker et al. [19] provided a systematic literature review (SLR) regarding "the rising role of big data analytics and IoT in Disaster Management." In 2017, Alam et al [12] presented an analysis and fusion of big data results acquired from IoT and related sources, which was intended to enable reliable, efficient, and accurate decision-making and management for environments. In the same year, Gu et al. [20] presented an exploratory study about IoT and big data as potential solutions to outstanding problems in the field of waste electrical and electronic equipment management (WEEE). From these results, it becomes clear that while some surveys do cover big data and IoT, such studies focus mainly on specific applications rather than general use or benefits. As a result, we observe that there is still a gap in the research for planning and designing integrated big data and IoT technologies in smart environment areas. Thus, to our best

knowledge, our work here the first systematic review of its kind concerning big data and IoT integration for smart environment applications.

Table 1: Related surveys.

Reference	Objectives and topics
Seker et al., 2019 [19]	It highlights the increasing role of big data analytics (BDA) and IoT in disaster management scenarios. According to an investigation of recent studies, a classification of a thematic taxonomy, a review of ubiquitous solutions, and a conceptual model for the deployment of IoT and BDA in disaster management all proposed.
Al Mamun et al., 2019 [21]	It presents a review of current research and developments on sensors and systems such as wearable devices designed for applications in environmental IoT. Descriptions and comparisons of state-of-the-art wearable environment monitoring-systems are provided.
Alam et al., 2017 [12]	It presents a literature review on data fusion for smart ubiquitous environment, focusing particularly on mathematically-derived methods and specific IoT environments (e.g. distributed, heterogeneous, ...). A list of the opportunities and challenges within each mathematical method and environment is also provided.
Martín et al., 2017 [20]	It reviews agro-industrial and environmental fields from an IoT basis, aiming to identify trends, areas of application , architectures, and open challenges present in these two specific fields.
Gu et al., 2017 [22]	It provides an exploratory study about the potential of combining big data and IoT to find new solutions to the management of waste electrical and electronic equipment (WEEE) problems. The combined applications brings a host list of new opportunities and challenges.

## 1.2. Contributions and research- structure

The contributions of this review are summarized as follows:

- The main research published recently regarding big data and IoT for smart environment applications are investigated
- The principal benefits and key requirements of big data and IoT for smart environmental applications are identified and discussed
- The main tools and technologies related to big data and IoT in smart environment approaches are identified
- A discussion about the main results and contributions of the previous research has been

conducted

- A set of open perspectives and future research directions are highlighted.

These contributions are presented as follows. In Section 2, Section 2 states the methodology of this research and our research questions are detailed. Results and answers to the questions raised in this research are detailed in section 3. Then, our conclusions and thoughts on directions for future work are presented in section 4.

## **2. Research Methodology**

For this paper, a systematic review (SR) is conducted using the Preferred reporting items for systematic reviews and meta-analysis (PRISMA) approach [23]. PRISMA is an evidence-based minimum set of items used to guide the development and structure of SRs and other meta-analyses. In our paper, the reviewing protocol includes three steps, namely: definition of research questions, search phase, and specification of inclusion and exclusion criterion. The specifics of these steps for our research study are described in the subsequent sections.

### *2.1. Definition of research questions*

This SR is laid out in such a way that covers the scope of research reviewed by categorizing and reviewing existing related publications. The first step consists of defining the research questions in order to precisely describe the coverage rate of existing works. By studying related works, we can provide several insights that can then help researchers develop new ideas [24].

The research questions used in our SR are described in Table 2.

Table 2: Research questions.

<b>Research question</b>	<b>Motivation</b>
RQ1. What is the distribution per year, authors, country, context, and publication venues of published papers related to IoT and big data in Environment applications?	The answer to this question allows identifying the when, where, and by whom the research studies have been conducted.
RQ2. What is the purpose of combining IoT and big data in environment applications?	The answer to this question illustrates the need for using IoT and big data in new smart environment applications.
RQ3. What are the main challenges facing IoT and big data in environment applications?	The answer to this question helps exploring the main difficulties that arise when using IoT and big data and depicts the advantages and limitations of existing solutions.
RQ4. What are the main tools and technologies related to big data and IoT in environment applications?	The answer to this question helps identify appropriate tools and techniques adopted by current applications.
RQ5. What main results and conclusions have been drawn based on big data and IoT in environment applications?	The answer to this question provides the main outcomes of the studied works.
RQ6. What are the future research directions and open perspectives of IoT and big data in environment applications?	The answer to this question helps identifying the unexplored relevant research ways.

## 2.2. Search phase

To conduct our SR, the first step is to define information sources. As depicted by the Table 3, various academic databases, digital libraries, and search engines – both academic and open access – have been searched. The next step consists of defining procedures for exploring the scientific and technical documentation that these searches returned, in order to find relevant papers to our context. The proposed procedure is based on two main steps: (i) determining search terms from the previous research questions in order to obtain a set of keywords; (ii) using Boolean operators, AND/OR, to determine queries that will be used to find and collect all related results .

Table 4 depicts the search queries used in this paper.

Table 3: Search sources

Source	Type	URL
Science Direct- Elsevier	Digital library	<a href="http://www.sciencedirect.com/">http://www.sciencedirect.com/</a>
Scopus	Search engine	<a href="http://www.scopus.com/">http://www.scopus.com/</a>
IEEE Xplore	Digital library	<a href="http://ieeexplore.ieee.org/Xplore/home.j">http://ieeexplore.ieee.org/Xplore/home.j</a>
ACM Digital library	Digital library	<a href="http://dl.acm.org/dl.cfm">http://dl.acm.org/dl.cfm</a>
Web of science	Search engine	<a href="https://www.webofknowledge.com/">https://www.webofknowledge.com/</a>
Wiley online library	Digital library	<a href="https://onlinelibrary.wiley.com/">https://onlinelibrary.wiley.com/</a>
Google scholar	Search engine	<a href="https://scholar.google.com/">https://scholar.google.com/</a>
Sensors	Digital library	<a href="http://www.mdpi.com/journal/sensors">http://www.mdpi.com/journal/sensors</a>
Springer	Digital library	<a href="https://www.springer.com/">https://www.springer.com/</a>
Researchgate	Social networking site	<a href="https://www.researchgate.net/">https://www.researchgate.net/</a>
Edinburgh library database	Digital library	<a href="https://my.napier.ac.uk/Library/">https://my.napier.ac.uk/Library/</a>

Table 4: Search queries

TITLE-ABS-KEY
S1 ((big data AND Internet of Things) OR (big data AND IoT))
S2 ((big data AND environment applications)
S3 ((environmental applications AND Internet of Things) OR (environment applications AND IoT))
S4 ((big data AND Internet of Things AND environment applications) OR (big data AND IoT AND environment applications))

### 2.3. Inclusion and exclusion criterion

In order to refine search results, we use a set of inclusion criterion (IC) and exclusion criterion (EC) to determine relevant papers (Table 5). Studies that fail to answer EC are ignored; in addition, a screening process is applied to select relevant papers to our context. The screening process is based on three IC steps:

**(i) Abstract-based step:** we discard irrelevant results based on information and keywords found in paper abstracts. Papers whose abstracts satisfied at least 40% of IC were kept for further processing.

**(ii) Full-text-based step:** we discard results that did not address or refer to the research terms in the Table 3, i.e. papers that only represent minor aspects of the search terms represented in their abstracts.

**(iii) Quality-analysis-based step:** we applied a quality analysis to the remaining results by removing those that did not satisfy any of the following criteria:

–**C1:** The paper discusses a comprehensive solution of big data and IoT in environment applications.

- C2:** The paper includes the technical implementation of the proposed solution.
- C3:** The paper includes related works.
- C4:** The paper presents a discussion of the obtained results.

Table 5: List of the IC and EC

<b>Inclusion criteria</b>
Studies should be listed at least in one of the research databases
Studies are published during the period between 2011 and 2019
Studies should meet at least one of the search term
Studies should be published/in-press at a journal, conference, and magazines
Studies should provide answers to the research questions
The search is performed based on the title, abstract, and full text
<b>Exclusion criteria</b>
Studies that are not written in English
Duplicated papers
Studies with missing full text
Papers not directly relevant to IoT and big data in environmental applications

#### 2.4. Conduction of SR

The selection of relevant articles to this review was done through the following key steps: identification, screening, eligibility, and inclusion as detailed by the PRISMA flow diagram. Figure 4 illustrates the systematic search-strategy process. Initially, we obtained 240 records; 212 identified through electronic databases and research engines (Table 3) and 28 additional records identified through other sources. These results were reduced to 173 after excluding duplicates. Then, eligibility criteria based on the title and abstract were applied to these 173 and this elimination round reduced our results to 50; further eligibility criteria based on the full text finally enabled us to obtain 45 results. These 45 papers will be analyzed in depth to extract the results presented in the next section.

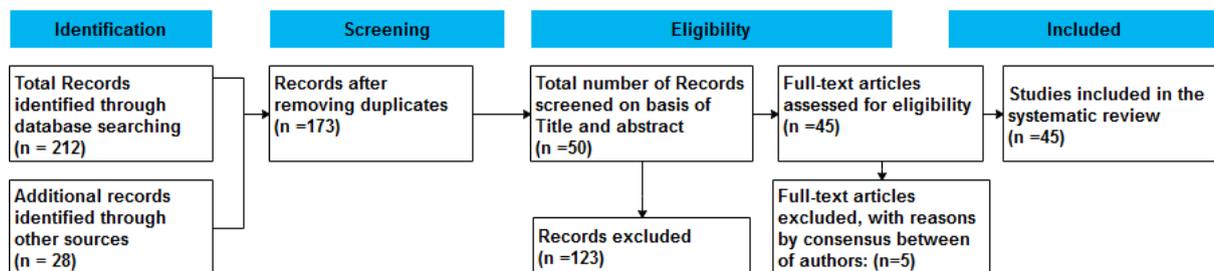


Figure 4: The systematic review process

### 3. Results

#### 3.1. Answers to the research questions

Following our review and analysis of the 45 selected articles, this section details the results of our SR and provides answers to our research questions discussed above based on the findings of this process.

*Answer to research question RQ1 : **what is the distribution per year, authors, country, context, and types of publication of published papers related to IoT and big data in environment applications?***

We have selected only papers with relation to environment applications such as agriculture, water, oil, soil, air, urban, and transportation from all the different studies that we found. The total number of publications is 45. Figure 5 depicts the distribution of selected papers by (a) publication year, (b) publisher, and (c) domain. The diagram in Figure 5 (a) demonstrates that the trend of using big data and IoT in environment applications is increasing, and that the combination of these two technologies had not been proposed prior to 2013. However, after 2013, this trend ascended, with 20 important papers published in 2017 and 12 more in 2018. Furthermore, most of the selected papers are published in IEEE Explorer with 32 papers followed by Elsevier with 11 papers. Figure 5 (c) demonstrates that most selected papers are related to disaster management with 17 papers in this area, followed by smart agriculture with 9 papers. Figure 5 (d) then reveals the country of origin for all papers selected. The United States has contributed to the largest amount of papers for a single country, while Asia in its entirety as a continent has contributed twice this number. It is also worth pointing out that both North Africa and the United Kingdom have contributed the same number of papers.

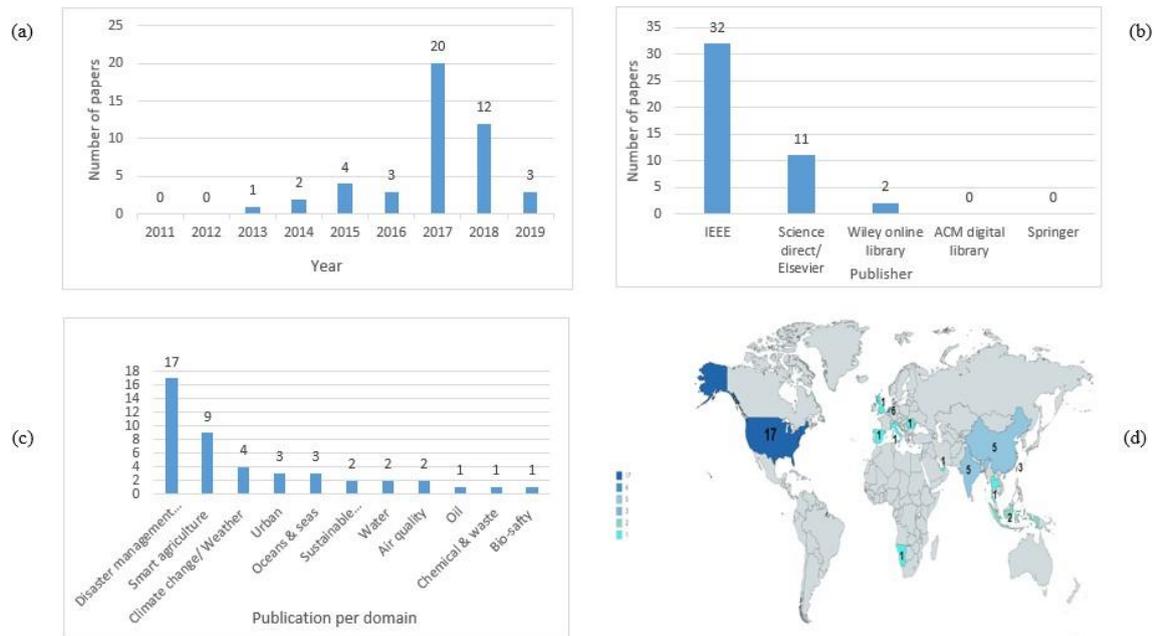


Figure 5: Distribution of selected papers by (a) publication year, (b) publisher, (c) domain, and (d) country

*Answer to research question RQ2 : what is the purpose of combing IoT and big data in environmental applications?*

Big data tools are often used to store massive volumes and varieties of collected data, to run parallel and real/quasi-real time analysis on that data, and to employ data mining techniques to extract information from it. Therefore, big data technologies and IoT can lead to many interesting results in smart environment applications. Some of the main benefits are as follows:

**1- Multi-source and heterogeneous data**

By combining IoT infrastructure equipped with various data sources (e.g. sensors, remote sensing, cameras, and smartphones) with big data technologies assisting in data management (e.g. data collection, data storing, data processing, and data analysis), a significant variety of data sources can be integrated with one another to deliver useful new information and insights. Data can be collected from IoT sensors as well as other sources to enable the construction of predictive models [25]. Therefore, with the growing availability of rich new data sources, using big data and IoT can provide interesting results in smart environment applications.

**2- Connectivity**

Connectivity is the most important task required by big data and IoT, as it facilitates the aggregation of massive volumes of heterogeneous data from various sources. In this way,

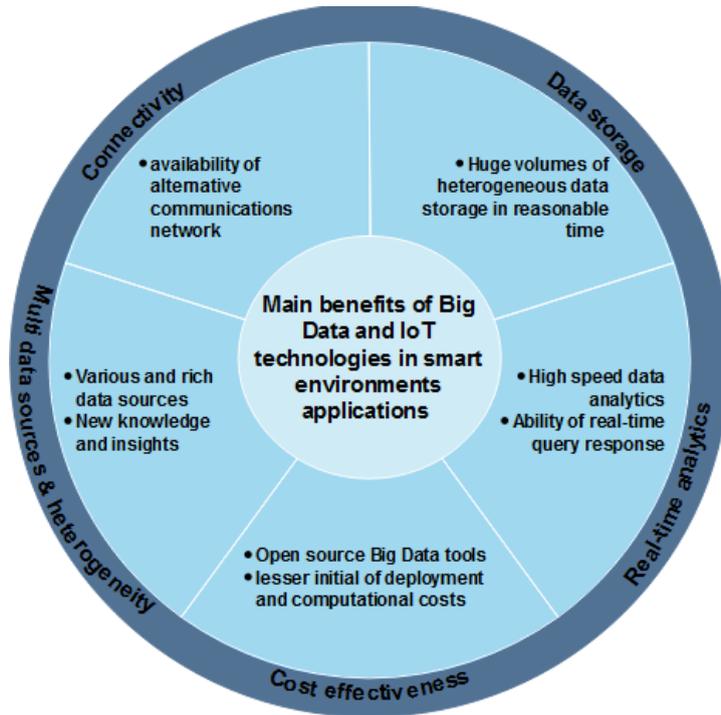


Figure 6: Main benefits of big data and IoT based smart environment field

connectivity is required in order to ensure high-performance computing as well as all information exchanges among devices and environment management authorities [26]. One of the major purposes of combining big data and IoT is the growing availability of different communication technologies that helps dealing with a large amount of data existing in smart environment applications. The connectivity between interlinked data nodes and the data management systems (DMS) serves as the backbone that ensures operational success. Therefore, because many types of communication technologies are now available, the proposed architectures in smart environment applications must be flexible in order to deal with distinct communication protocols, both remote and local.

### 3- Data storage

The storage of massive volumes of heterogeneous and unstructured data collected from different sources in real/quasi-real time is often challenging or even impossible with traditional DMS. In an environmental field characterized by this type of data, big data technologies such as Hadoop/Mapreduce, Hadoop Distributed File System (HDFS), Spark, Storm, Kafka, Hive, and NoSQL databases (e.g. Redis, Hbase, and MongoDB) have been developed to facilitate the collection, storage, and preservation of greater data volumes and more heterogeneous data sources [27]. When using IoT devices and similar data sources, having an infrastructure with effective storage capabilities can increase the efficiency and efficacy of data processing while also improving application design and function. Moreover, based on

BDA (big data analysis) technologies, it becomes possible to enable low-latency efficient-processing for data analysis while also maintaining the necessary storage structures for large datasets on low-cost commodity hardware.

#### **4- Data analysis**

In many environmental applications, such as disaster management, early warning systems based on real-time analysis are a key requirement, but one that comes with challenges. Connectivity between different data sources often results in the high-speed generation of huge volumes of data, creating a dynamic and demanding situation that challenges system performance of real-time data analysis [28]. To cope with this challenge and enable real-time processing, software solutions and technological capabilities dedicated to these processes are needed. Big data today makes such solutions possible, which can be crucial in decision-making for timely, effective emergency applications [29] [30].

#### **5- Cost effectiveness**

Many big data technologies are open source, which offers an important reduction in the cost required for the development and deployment of new applications. Cost-effectiveness is a particularly crucial in developing countries, where the lack of sufficient funds may prevent environmental management systems from being deployed at all [31]. Thus, developed more open architectures and platforms for big data and IoT could increase their adoption in these countries, facilitating the deployment of IoT in specific rural contexts that would greatly benefit from them (e.g., agriculture and climate change in Sub-Saharan Africa).

#### **Answer to research question RQ3: what are the main challenges of IoT and big data in environment applications?**

The articles surveyed here listed several key challenges [32] [33]. A list of the main ones are listed as follows, and can be divided into three parts according to [34]:

**1- Business:** IoT may be split into three areas based on its usage and users ;

- **Industrial IoT:** this subgroup is related to industrial-based applications such as wastewater systems, robotic manufacturing, and other industry-related applications of connected devices and systems.
- **Commercial IoT:** this subgroup comprises items used for business purposes, which may include devices trackers, medical devices, and inventory controls.
- **Consumer IoT:** this subgroup includes all IoT used by individual consumers for personal or leisure purposes, and may include devices such as phones, computers, smart cars, and entertainment systems.

**2-Society:** understanding IoT from the consumer's the point of view is complicated because consumers are not experts in this technology and their requirements are dynamic , often changing with the release of new technologies.

**3- Technology:** This includes all technologies required to generate functionalities related to IoT systems, both as a standalone solution and as a part of current systems.

**3-1 Connectivity:** connecting and maintaining stable connections among multiple devices will be one of the largest and also most important challenges in the future of IoT, and it is predicted to defy the existing structure of current communication models. At present, a centralized server/client paradigm is being used for authenticating, authorizing, and connecting several nodes across a network.

**3-2 Security:** IoT has already caused serious security issues that have troubled different public and private sector companies across the world. The massive number of new hubs being added to the networks and the internet will provide attackers with innumerable attack vectors to invade the system, especially since a noticeable number of them suffer from security deficiencies. Hacking cameras, for example, and different monitoring devices and systems are signifying a security nightmare being caused by the future of IoT.

**3-3 Compatibility and durability:** IoT's rapid expansion has led to various technologies competing to become a standard, which may cause future problems ensuring compatibility among them. Issues of compatibility will require the development and deployment of additional hardware and software to ensure that devices from different manufacturers and industries can be interconnected.

At the same time, though, some of these technologies will ultimately become obsolete in the next few years. This is very important since IoT devices tend to stand in service for many years, compared to generic computing devices (lifespan of a few years), so they should be able to function even when their producer goes out of services.

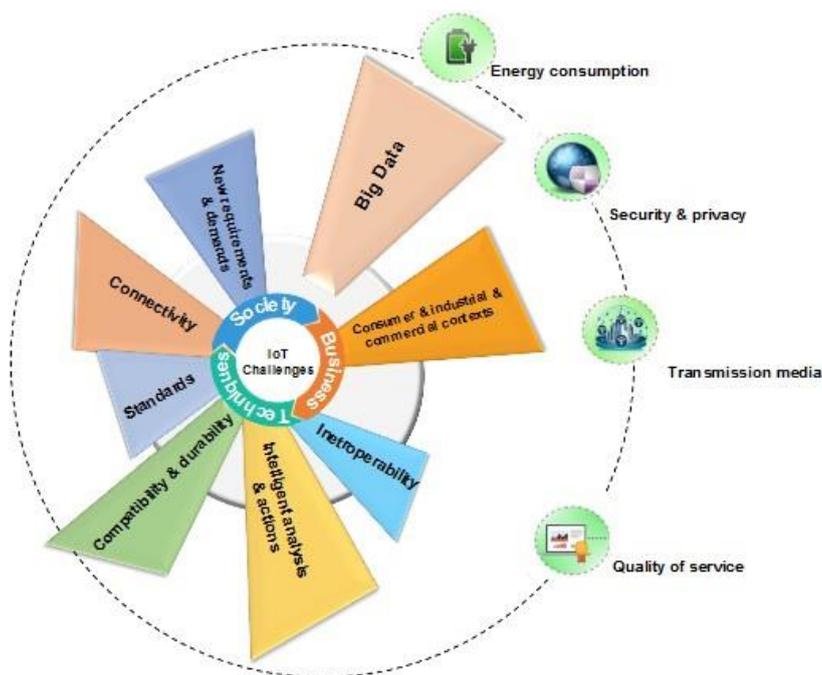


Figure 7: IoT challenges

**3-4 Standards:** : this is the sum of operations required to build an IoT infrastructure, as comprised of network protocols, data aggregation, and communication protocols. Two main issues face the adoption of standards in IoT:

- Standards for managing unstructured data.
- Technical skills required for leveraging new tools for data collection.

**3-5 Smart analysis and actions:** the last phase in the implementation of IoT concerns the extraction of knowledge and insights for analysis, as driven by suitable services for accompanying models that assist in using cognitive technologies. Sufficient services for analysis should be automatically identified, especially in dynamic environments, and this requires convenient semantic means of describing their functionalities. Challenges facing the adoption of smart operations in IoT include:

- The security and privacy of information
- The influence of machines on human behavior
- Unpredictable conditions of machines operations
- The interoperability of machines
- The moderate adoption of recent technologies
- The inaccurate analysis because of the uncertainty in the data and/or models
- The inability of traditional systems to analyze unstructured data
- The inability of traditional systems to manage streaming data

Table 6 depicts these challenges .

The various challenges of IoT and big data in environmental application fields can be summarized as depicted in the Figure 8:

**Volume:** describes the amount of data and information that we have. This quantity was measured in Gigabyte (GB) size, then exponentially moved to zettabytes (ZB) and yottabytes (YB). Iot is making an expressive increase in data. This phenomenon presents the heart of what big data means, which is the requirement of many operational mechanisms in addition to new technologies for data storing, processing, and management.

**Velocity:** means the speed of data access and generation. How can we manage these data in real/near-real time is one of the big issues of IoT and big data?

**Variety:** is a key challenge facing IoT and big data in general, and it becomes especially relevant in the field of environmental applications. Data can be structured (e.g. text), semi-structured (e.g. XML and JSON document), or unstructured (e.g. image or video), and it is not an easy assignment to organize such disparate types of data.

**Veracity:** specifies the level of quality which means how accurate and applicable is the data. In other words, it is about the noise, biases, and abnormalities in data. Is this data relevant and meaningful to the problem analysis [35]. Veracity in data analysis is the most crucial challenge when compared to the other V's (i.e. as volume, variety, and velocity).



Figure 8: Big data challenges

**Value:** refers to the worth of the data being extracted. It is important to understand the costs and benefits that can be obtained after collecting and analyzing data characterized by its volume, velocity, and variety which takes both time and resources.

**Visualization:** it refers to the implementation of new visualization techniques (e.g., graphs and charts) to effectively highlight the relationships within data in a better manner.

**Security and privacy:** the challenge of privacy and security is a formidable obstacle for IoT, as the risk of an attack and hacking of confidential information is very common due to the diversity of used devices. Moreover, additional requirements would also be important in the IoT, besides the protection and security (e.g. trustworthiness and authenticity of partners in a communication). This issue can be divided into three main considerations:

- Vulnerability to hacking.
- Trust consideration.
- Data protection.

**Interoperability:** denotes the ability to combine, link, and manage at least two datasets. Different types of data collected from various sources may not present obvious relationships with one another and can also be difficult to align. Therefore, it is a very important and challenging task to attain the maximum level of interoperability during both the real-time collection and the later integration of diverse data.

**Data cleansing:** refers to making data usable after having collected it in its raw form. This task is particularly important for smart environmental applications, where incomplete or ambiguous data can lead to incorrect analyses, wasting time and resources while delaying

Table 6: The main challenges facing big data and IoT for smart environmental applications

Challenges of Big data and IoT	References
1- Data volume	[37] [38] [28]
2- Data value	[19] [37] [25] [39] [40] [41]
3- Data variety (cleansing, normalization, integration)	[19] [42] [43] [40] [41]
4- Data veracity and uncertainty modeling	[42] [41] [19]
5- Data storage	[19] [38] [40]
6- Data analysis	[38] [41]
7- Real-time data analytic	[19] [42] [37] [28] [40] [41]
8- Data visualization	[25] [28]
9- Security and privacy	[42] [25] [28] [39] [43] [44]
10- Communication and interoperability	[19] [37] [39] [41]
11- Complexity of architecture (connectivity, deployment, energy consumption, fault tolerance)	[25] [28] [39] [41]

action. Therefore, data cleansing restrictions will determine the efficiency of the analysis, achieved in a particular dataset [36]. At the same time, though, harmony should be kept within the model of data cleansing and the accuracy enhancement of analysis. Indeed, data cleansing requires complex relationship models, which often necessitates additional power for computation and time for processing.

**Real-time data analysis:** denotes the analysis of data immediately once it becomes available, including the processing of shifting data sources and handling large amounts of data to return answers and insights in less time.

**Knowledge extraction:** refers to the ability to extract knowledge from heterogeneous data gathered from different sources with various formats and types, so that the data can be more productive and effectively stored.

*Answer to research question RQ4 : What are the most popular tools and technologies related to big data and IoT in smart environmental applications?*

After examining the use of IoT and big data in different environment applications, we noticed that Hadoop was the most commonly used ecosystem in these studies. However, other services may be used as well.

**Apache Hadoop** this is an open source project of the Apache Foundation, which serves as the spine of any big data architecture. This framework is intended for distributed storage and distributed processing of massive amounts of data on computer clusters built from commodity hardware, which is a much cheaper and more effective solution than running a dedicated data center. The ecosystem of Hadoop includes different frameworks and tools to handle various types of storing, data aggregation, processing resource management, searching, and data analysis for different types of heterogeneous data (structured, semi-structured, and unstructured) [37]. The main core of Apache Hadoop consists of two sub-projects, which are HDFS for storage, and MapReduce for processing. First, Hadoop breaks a large task

into numerous task blocks and distributes them across different nodes in the cluster [19]. MapReduce then transfers packaged code to the nodes to be processed in parallel. As long as the previous process does not demand similar nodes, each block of data can be processed in a different separate node [31]. Then the fragmented processing outputs are combined to form the final processing result. Apache Hadoop is a highly reliable and efficient framework, and as such, it is one of the best solutions for developing applications capable of execution on clusters of computers. Moreover, it can perform complete statistical analysis for large volumes of data [45].

**Apache Flume:** this is an efficient distributed service for collecting, transferring and aggregating large volumes of data streams such as log file and event into Hadoop [46]. The flume is a reliable, available, distributed, and configurable service that can be used to collect data from many real-time servers. The main use-case of this service is to ingest data into Hadoop.

**Apache Kafka:** this is a distributed, high-throughput, publish-subscribe messaging platform. It is the most common way to get data in Apache Storm [47]. In contrast with other traditional systems, Kafka has better performance, unrepeatable and unchangeable classification and high failure tolerance. More than that, it is also suitable for message processing on a large scale and enables data to be used by various applications.

**Apache Sqoop:** this is an open-source tool offering a parallel load of data [48]. The sqoop is used to transfer bulk data between structured data stores (i.e., relational databases) into HDFS and vice versa.

**Apache Spark:** this is an open-source fast-clustering computing engine designed for high-speed computing of large-scale data. Spark is developed on Hadoop capabilities but is claimed to be faster because of its in-memory capability, interactive queries, and stream processing [19]. Spark adds more concepts such as low latency high-level APIs in various languages such as Java, Scala, Python and R, as well as a stack of high-level tools. Also, it supports an algorithm library about big data machine-learning algorithms like clustering, classification, and regression, deep learning such as MLib (Machine Learning Library) [49]. Spark is mainly used to increase the processing speed of the application and it can be used with Hadoop data sources as a programming model for processing [37].

**Apache Storm:** this is a robust distributed framework for real-time computation. Like Hadoop, Storm is an open-source framework mainly used to analyze massive data [48]. Both frameworks complement each other, yet they differ in some respects, such as the way in which Storm is not stable but it can execute all the functions and Hadoop is good with MapReduce yet it encounters delays in real-time calculations. Storm is able to process a big number of records per node per second on a cluster of moderate size and it supports different programming languages like Scala, Python and Scala [50].

**Apache Hive:** this is an open-source data warehouse project that runs on top of Hadoop. Hive uses an SQL-based interface to inquire about data stored in different file systems and databases that coordinate with Hadoop [41]. Hive is a fast, accessible way of responding to complex queries through SQL-like queries (e.g. HiveQL) and to analyse big data [37].

**NoSQL database:** this technology's name refers to Not Only SQL database. NoSQL is used to manage massive amounts of unstructured and semi-structured data perfectly, in contrast with traditional relational databases, which are not suitable for these data types. In other words, it is managed by a non-relational DMS and does not require a fixed schema (i.e. schema-less). NoSQL databases provide four different data model storage, each one of which has its own advantages and limitations (Key-Value, column data model, document data model, and graph data model) [53]. There is no one solution that is better than the others, but there are some databases that are more suitable to solve specific issues. For instance, MongoDB, the document-oriented database, is commonly used in studies on smart agricultural applications and smart environmental monitoring [25] [48] [52] [50] [51]. MongoDB prominent among NoSQLs for its speed, scalability, efficiency, and fluency in the query language [54].

**Cloud computing:** this is a kind of computing, which relies on shared remote server presented on the internet for data management, storing, and processing data, instead of using personal hosts or local servers to handle applications [25]. As many smart environment applications need to analyze large volumes of streaming and historical data with a low-latency time of data processing, it is an appropriate belief to suggest the cloud computing technologies to increase the speed of data analysis [42]. Big data and cloud computing technologies can be treated as "one body and two sides" in which cloud computing ensures the basic layer of computing resources and backing the higher layer of big data processing, while big data provides the application platform for driving cloud computing (e.g. Hadoop). Both IoT and cloud computing have seen a fast and independent progression. Although these two technologies are distinct, many aspects are complement one another, so that IoT can profit from the resources offered by cloud computing and the virtually unlimited capabilities to reward its technological restraints (i.e. storage, processing, and communication) [31]. There are three main services the cloud offers to respond to user needs, which are; Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) [55]. Finally, there are the three categories of cloud computing (i.e. private, public, and hybrid), correspond to different kinds of user demands. The most common platforms used by cloud computing in environment applications are shown in Figure 6.

An overview of the main tools that our 45 selected articles used regarding IoT and big data in smart environment applications areas is shown in Figure 9.

**Access technologies:** a robust, reliable, and energy-efficient data transmission acts as the backbone for any smart environment system. From a communication perspective regarding any smart environment application, access technologies should ensure stable network connectivity and optimized services for effective data transmission between data collection sensors and back-end servers. It is a crucial matter to ensure the flow of data, the connectivity the network, and the safety of all components, especially for warning cases (e.g. environment monitoring) or emergency events (e.g. disaster management). Therefore, a combination of different communication protocols and networks is required to provide whole network infrastructure for data transportation and to help connect diverse data sources.

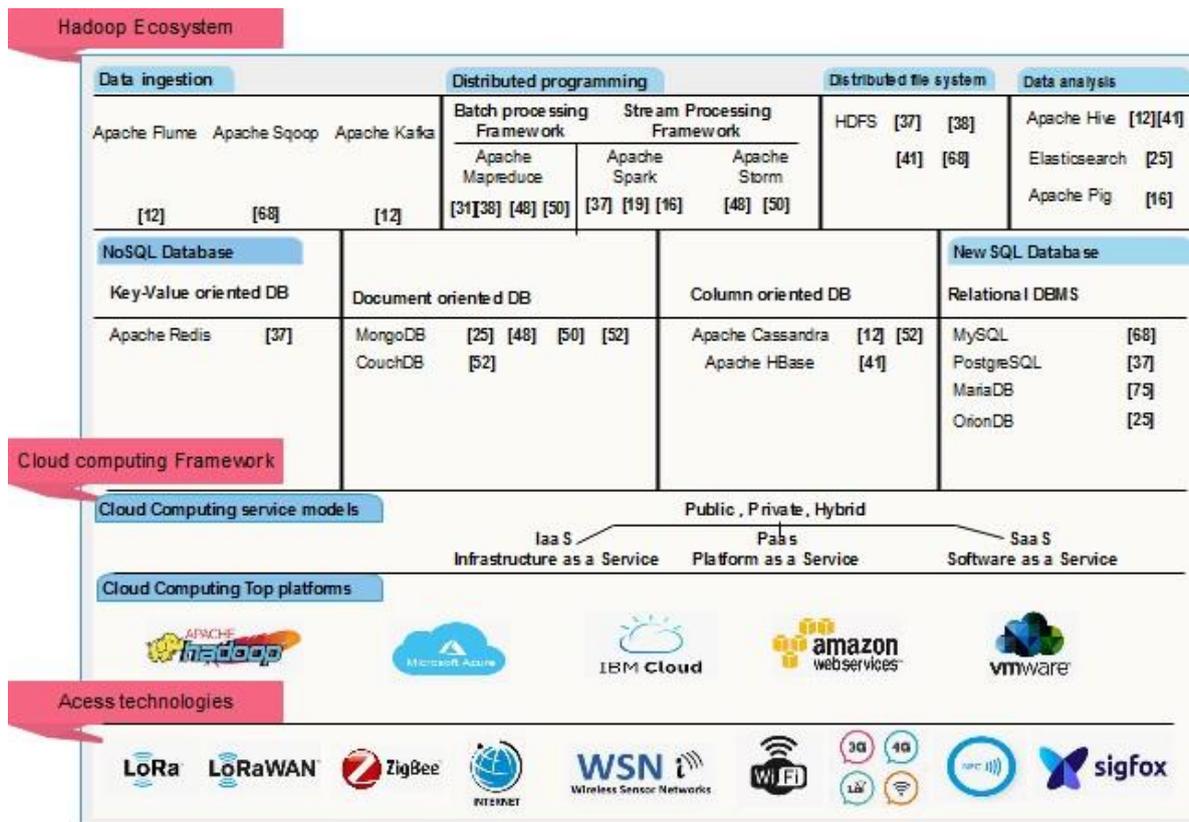


Figure 9: Common tools and technologies in big data and IoT for smart environmental applications

Table 7 classifies the papers reviewed based on different access technologies and IoT protocols appropriate in smart environment applications. Some of the main access technologies suitable for smart environment communication are 4G/LTE (Long Term Evolution, also called 4G), ZigBee, Wi-Fi, Bluetooth, radio frequency identification (RFID), general packet radio service (GPRS), wireless sensor network (WSN), long-range (LoRa), long-range wide area network (LoRaWAN) [56] [57] .

ZigBee is a low-data rate, low-power, and close-proximity wireless ad-hoc network. This technology is intended to be simpler and less expensive than other wireless personal area networks (WPANs), such as Bluetooth or more general wireless networking such as Wi-Fi. RFID is a kind of wireless communication that integrates the adoption of electrostatic or electromagnetic combination in the radio frequency portion of the electromagnetic spectrum. WSN indicates a collection of spatially separated and dedicated sensors for recording and monitoring the physical conditions of the environment and coordination of data gathered at a central location.

WSN measures the physical environment (e.g. humidity, sound, pollution levels, temperature, and wind) [58]. LoRa defines the lower physical layer, while LoRaWAN was developed to define the upper layers of the network [59]. With the evolution of IoT and LoRa technology,

WSNs become quicker, more reliable, and more robust [60]-[61]. GPRS is a packet-oriented mobile data standard of cellular communication network (i.e., 2G and 3G) of the global system for mobile communications (i.e., GSM). According to the studies examined, short-range communication technologies (e.g. ZigBee, Bluetooth, and RFID) are more popular than long-range technologies (e.g. cellular communication systems). Therefore, they could be a prudent solution for nearly all the environmental applications, since long-range communication technologies have many prevalent restrictions, including the high cost of initial investment, the limited number of connections, and the expensive prices for end-users (especially in rural places and countries such as Africa and India). On the other hand, short communication hardware devices are smaller, cheaper, and easier to set up.

According to Table 7, the main important technologies in IoT architectures in smart environment applications were short-range communication technologies such as 4G/LTE (72.72%), Wi-Fi (50%), WSN (45.45%), and ZigBee (40.90%). In addition, some protocols were used to enable the connection of sensor networks with the internet and has been designed to ensure the interoperability of sensor networks and the internet, such as LPWAN (low-power wide-area network), MQTT (Message Queuing Telemetry Transport), LAN (Local Area Network) and COAP (Constrained Application Protocol)/UDP (User Datagram Protocol).

*Answer to research question RQ5 : **What main results and conclusions have been drawn based on big data and IoT in environment applications?***

To answer this question, our 45 selected articles are evaluated according to the application domain, data source, data type, data processing type, architecture details, and main findings, as shown in Table 8.

We note that most big data and IoT-based applications in smart environmental fields are applied in a real-time case to multi-source datasets collected from sensors (e.g. water sensors, agriculture sensors, weather sensors, and air quality sensors). Most architectures reviewed in the literature used the following steps:

- Data acquisition step based on IoT infrastructure for collecting data from sensors
- Network layer for information transmission and interconnection of the platform
- Preprocessing of data including cleaning to remove noise and redundancy, instance selection, normalization, transformation, feature extraction, and selection, etc
- Big data architecture to store and process a huge volume of data
- Data analysis including many complex tasks such as descriptive, predictive, prescriptive, and statistical
- Presentation and data visualization responsible for finding useful insights to interpret results and to communicate information to users

Table 7: Comparison of existing access technologies

No.	Ref.	LoRa/LoRaWan	ZigBee	Wi-Fi	GPRS	Bluetooth	4G/LTE	RFID	WSN	IoT protocol
1	[42]			*		*				-
2	[19]		*	*	*	*	*		*	-
3	[25]	*					*		*	-
4	[64]						*			-
6	[65]						*		*	-
7	[66]			*			*		*	-
8	[43]		*			*	*			-
9	[38]		*						*	-
10	[67]		*			*		*		-
11	[68]		*	*		*	*			-
12	[48]		*	*		*	*			-
13	[50]			*						MQTT
14	[69]			*			*			-
15	[39]	*	*	*		*	*			LPWAN
16	[70]			*			*			-
17	[71]			*	*					-
18	[22]							*	*	-
19	[72]								*	-
20	[55]				*		*		*	LAN
21	[73]		*		*				*	-
22	[74]	*	*	*			*		*	COAP/UDP

Another important observation is that most proposed applications in the smart environment field use the cloud to store large amounts of heterogeneous data being collected by IoT-based sensors. The processing of big data is achieved using big data tools such as HDFS, MapReduce, Hive, Spark, and Redis, which provide more flexibility and better execution time. Big data analytics in IoT involve streaming and storing a large volume of data. Providing real-time services and monitoring in environment applications requires the use of high-performance computing (HPC) solutions. Through many experiments, we can conclude that the combination of IoT and big data can provide new opportunities and applications in many environment-related sectors. Real-time queries can help organizations gain useful insights, make appropriate decisions, and interact with people in real time. Results have great social and economic impacts in many fields, including water contamination and preservation, crop management, precise agriculture, control of agricultural product-costs, weather forecasting, and air quality, among others [63] [62].

Table 8: Comparison of main results in considered studies.

Ref	Application domain	Data source & Data type	Data processing type	Architecture details	Main findings
[66]	Water supply	Data collected from water sensors	Real-time	Four main components: - Water sensors to provide pressure and flow data - GIS - Big data to store and process water data - IoT infrastructure for collecting data and delivering information to customers	+ Proposition of an approach based on big data and IoT to automate leak detection, to optimize water supply, production, and energy consumption and to optimize water consumption. + Combination of IoT and big data to develop a system that allows sustainable water supply.
[37]	Water monitoring	Data collected from water sensors	Real-time	Five main components: - Remotely operated underwater vehicle (ROV) equipped with water quality sensors - Portable water quality unit to collect and save the data - Coral reef to prevent coral bleaching - Big data for processing huge volume of data - Wireless mesh network access to ensure the connection between ROVs.	+ Proposition of an open framework that combines big data and IoT for underwater monitoring + Use of big data tools (e.g. HDFS, MapReduce, Hive, Spark, and Redis) provides more flexibility and better execution-time compared to relational database management systems, especially when working with larger volumes of heterogeneous data.

Table 8: Comparison of main results in considered studies.

[25]	Three use cases: fish farming, cattle rustling, agriculture weather information	Data is collected from IoT sensors	Real-time	<p>Three main components:</p> <ul style="list-style-type: none"> <li>- Platform as a service (PaaS) and IoT approach to reduce development costs</li> <li>- Big data techniques to process the big volume of data produced by sensors</li> <li>- Local and global cloud to deal with the connection disconnections</li> </ul>	<ul style="list-style-type: none"> <li>+ Open architecture and platform for IoT and big data.</li> <li>+ Addressing the intermittent of the Internet connection with local and global cloud and IoT services at the edge.</li> <li>+ Dedicated application to providing actionable information.</li> </ul>
[38]	Smart air-quality monitoring and its impact on brain health	Public data from the environmental protection monitoring and distribution center	Real-time	<p>The proposed approach is based on three main steps:</p> <ul style="list-style-type: none"> <li>- Big data storage and processing using MySQL for local data and the HDFS system for cloud storage</li> <li>- Analysis of data by proposing a brain-health quality-index prediction using Bayesian classification.</li> <li>- Brain-health quality monitoring based on IoT. + Proposition of brain health quality-index forecasting based on popular information technology, and an empty space based on BDA.</li> </ul>	<ul style="list-style-type: none"> <li>+ The prediction model of gas quality based on big data, IoT, and smart devices.</li> <li>+ Development of a cloud computing-based technology to enhance data processing.</li> </ul>

Table 8: Comparison of main results in considered studies.

[75]	Climate forecast and air quality simulation	Data are collected from air quality monitoring stations	Real-time	<p>The proposed application is based on:</p> <ul style="list-style-type: none"> <li>- The master-server node retrieves metadata from the database and distributes the information to slave nodes based on client requests</li> <li>- The slave data-node retrieves data by querying local databases</li> <li>- Data are partitioned according to time into different local databases</li> <li>- Communication between master and slave are achieved using a message passing interface</li> <li>- Results are displayed to the user through a web-based interface</li> </ul>	<ul style="list-style-type: none"> <li>+ Proposition of an experimental big data platform based on the HPC distributed computing and in-place computing.</li> <li>+ Use of big data platform for environmental monitoring to visualize global climate forecasting and air quality simulation.</li> <li>+ Providing real-time information services and monitoring related to the environment requires the establishment of standardized procedures to automate the processing of data.</li> <li>+ Two case studies related to global climate data visualization and air quality model simulation, and another application to identify countermeasures for reducing air.</li> </ul>
------	---	---	-----------	---	---

Table 8: Comparison of main results in considered studies.

[76]	Weather analysis	Data collected from 8000 weather sensors	Batch	<p>Five main components:</p> <ul style="list-style-type: none"> <li>- Data acquisition layer to collect data from sensors</li> <li>- ETL layer to parse and create semantically annotated data</li> <li>- Learning layer for semantic rule reasoning and feature extraction</li> <li>- NoSql database for data storage</li> <li>- Action layer to evaluate results.</li> </ul>	<ul style="list-style-type: none"> <li>+ Weather analysis using an IoT Framework based on data retrieval, data processing, and learning</li> <li>+ Fault-sensor detection using clustering analysis technique</li> <li>+ Implementation details of each component of the proposed framework</li> </ul>
[31]	Precision agriculture	Agricultural data coming from IoT devices	Real-time	<p>Two main components:</p> <ul style="list-style-type: none"> <li>- Cloud database to store agriculture data.</li> <li>- MapReduce to process and analyze data.</li> </ul>	<ul style="list-style-type: none"> <li>+ Decision support for crop sequence based on acquired data.</li> <li>+ Control of agricultural product-costs through continuous notifications to farmers about agricultural schemes.</li> </ul>

Table 8: Comparison of main results in considered studies.

[28]	Smart flood monitoring	Weather data	Real-time	<p>Architecture based on four layers:</p> <ul style="list-style-type: none"> <li>- IoT layer to collect information using IoT</li> <li>- Fog layer to reduce the latency of the overall application using the computation of network devices</li> <li>- Data analysis layer responsible for receiving, storing, and analyzing data. It is composed of four components; data preprocessing, dimension reduction, cluster analysis, and Holt-Winter's forecasting</li> <li>- Presentation layer responsible for the communication of information to users using a mobile application.</li> </ul>	<ul style="list-style-type: none"> <li>+ Proposition of industrial internet of things (IIoT) and big data-HPC to monitor and predict floods.</li> <li>+ Representation of geographical locations according to hexagonal structures.</li> <li>+ Dimensionality reduction of big data using singular value decomposition.</li> <li>+ Categorization of flood perception into five different levels using K-means.</li> <li>+ Prediction of future flood-rating using Holt-Winter's forecasting</li> </ul>
[64]	Agriculture	Data related to crop monitoring	Real-time	<p>An approach based on:</p> <ul style="list-style-type: none"> <li>- Cloud computing</li> <li>- Big data</li> <li>- IoT</li> <li>- Wireless sensor network to connect different components.</li> </ul>	<ul style="list-style-type: none"> <li>+ Cloud-based solution providing several services to farmers: crop management, finance management, market-oriented services, exchange of information, and farming monitoring.</li> <li>+ Communication through internet, SMS, voice, and a web portal</li> </ul>

Table 8: Comparison of main results in considered studies.

[69]	Urban Health-care	<ol style="list-style-type: none"> <li>1. Multidimensional air quality indicator</li> <li>2. Physiological big data</li> </ol>	Real-time	<p>The proposed system is based on :</p> <ul style="list-style-type: none"> <li>- Data collection</li> <li>- Data fusion at the edge-cloud-based level</li> <li>- Data integration on a remote cloud or meteorological super-computing platform</li> <li>- A healthcare monitoring system via urban big data + The proposed system analyzes data to improve the quality of life of urban residents</li> </ul>	<ul style="list-style-type: none"> <li>+ Integration of meteorological data, mobile device crowdsourced data, and IoT sensing data.</li> <li>+ A test is established to depict applications of the proposed system.</li> </ul>
------	-------------------	--	-----------	---	--

Table 8: Comparison of main results in considered studies.

[77]	Safe waterway navigation	Data collected from vessels	Real-time	<p>Three main steps:</p> <ul style="list-style-type: none"> <li>- Data cleansing to remove noise and redundancy</li> <li>- Data mining, extraction of ship trajectories and prediction of ships' draft information.</li> <li>- Data extraction and fusion to link trajectories and draft information.</li> </ul>	<ul style="list-style-type: none"> <li>+ Proposition of an approach to construct a vessel-safety navigation map that contains appropriate channel water depth information.</li> <li>+ The proposed method is based on automated identification system (AIS) data.</li> <li>+ Extraction and processing of both static and dynamic information about vessels through waterways from big real-time AIS data.</li> <li>+ BP neural network is used in the prediction process and the Hermite interpolation is used to complete the data between two location points of vessel.</li> </ul>
------	--------------------------	-----------------------------	-----------	--	--

Table 8: Comparison of main results in considered studies.

[39]	Precision agriculture	Agricultural data coming from IoT devices	Real-time	<p>An approach based on Internet Of Underground Things (IOUT) composed of by :</p> <ul style="list-style-type: none"> <li>- Underground-embedded systems with communication and sensing components</li> <li>- Base stations that transfer collected data to the cloud</li> <li>- Mobile sinks installed in equipment that moves around the field periodically (e.g. tractors and irrigation systems).</li> <li>- Cloud services used for permanent storage and real-time processing of collected data.</li> </ul>	<ul style="list-style-type: none"> <li>+ In-situ sensing: accurate localized knowledge of the soil.</li> <li>+ Wireless communication: components are deployed on the field or within the soil.</li> <li>+ Inter-connection: devices are linked on a crop field to the cloud for seamless integration.</li> <li>+ Real-time decision-making: availability of decision about soil and crop conditions on real-time for managers.</li> <li>+Mobility: seamless support.</li> </ul>
------	-----------------------	---	-----------	---	--

*Answer to research question RQ6: **What are the future research directions and open perspectives of IoT and big data in environment applications?***

After having studied articles related to IoT and big data in environment applications, we now highlight main open issues and different ideas, which can be used for further study in these research areas in the future. The ideas and future research areas can be divided into six main areas as follows:

**Data collection**

In the fields of both IoT and big data, acquiring a huge volume of data from distributed sources is a challenging but necessary task for the implementation of big data and IoT technologies. Therefore, a protocol for data collection is critical, especially since this should be robust, secure, efficient, and energy-efficient. Among the studies that we had examined, only a few dealt with massive streaming data

collection. Thus, we identify the following important solutions for massive data collection of big data and IoT:

- Reducing volume by using methods and techniques of dimension reduction for data
- Storing data only after dimension reduction, not in raw form
- Using different methods, such as data mining to extract relevant features, before data storage

### **Data storage**

The storage of massive volumes of heterogeneous and unstructured data from different sources in real/quasi-real time can be challenging with traditional DMS, so having an environment with effective storage capability will increase the efficiency of data processing and can help to improve the applications so designed. The technologies of big data provide different tools to handle the large quantity of multi-source data, yet little of the research surveyed focused on distributed storage methods:

- Using more distributed storage tools, such as NoSQL databases and cloud-based storage, in the distributed environment will reduce the time needed for data analysis and decision-making later
- Data storage optimization

### **Real-time data analysis**

Real-time analysis is a key requirements of most environment applications, such as disaster management, early warning system, due to their dynamic and demanding nature [78]. In the majority of the papers we surveyed, the Hadoop/MapReduce was used for data analysis, yet none of these studies compared Hadoop/MapReduce, Spark, Storm, or Flink with other methods of data analysis for a specific application. Therefore, we propose the following solutions for big data and IoT:

- Choosing a particular analysis method based on a comparison between available methods or combining different analytical methods that decrease system workload will ensure on the efficiency of decision-making
- Developing new complex algorithms will reduce processing time while data volume increases.
- Reconstruction of novel and stable algorithms with a low degree of randomness.

- Exploiting deep learning (DL) techniques to extract insights from a large volume of [79] [80] .
- Identifying and analyzing which dataset can help to get effective real-time processing and to give correct results
- Including both batch and real-time big data processing mechanisms

### **Smart environmental architecture**

In general, we notice that the existing literature exhibits a lack of defined models for big data analysis and IoT for smart environment applications. Therefore, there is certainly a need for more detailed observations of various related reference models. We identify the following main needs:

- Proposition of novel architecture, which should be flexible and consistent in order to handle all the data sources and able to configure various network typologies for better data communication
- Designing a robust architecture to keep the environment resilient to handle any type of interruption caused by disasters, especially for applications regard disaster management, early warnings, and farming [81].

### **Standardization**

Standardization of IoT and big data is still in its inception. There are promoting ways about standards that include network protocols, data aggregation, and communication protocols, which are the sum of all the operations of data management. Therefore, standards such as communication protocols, security protocols, metadata, and data aggregation standards still need to be formalized.

### **Security and privacy**

IoT technologies have been extremely expanded in many fields. such as environment monitoring, agriculture, security, and emergencies (e.g. perimeter access control, liquid presence, radiation levels, and explosive/hazardous gases). Therefore, the security and privacy challenge is a major bottleneck for both big data and IoT domains, and the risks of online attacks and theft of confidential information is one of those issues. Moreover, many open-source tools and methods for big data analytics suffer from insufficient security mechanisms such as the Hadoop Ecosystem. Most of the papers we surveyed still lack security aspects in their platforms, particularly those related to agriculture [82]. Thus, important needs to be addressed in the future are as follows:

- Considering security as a required and high-priority task in different layers of IoT platforms

- Addressing the privacy needs in numerous aspects of IoT as required to protect the whole system from hackers, penetration, unauthorized entry, and destructive activities
- Using proper encryption methods to ensure the confidentiality of important data or data mining methods to analyze behavior patterns
- Use advanced security mechanism and new services to ensure the privacy of communication, data anonymity, and the transfer of information

#### 4. Conclusions

This paper reviews big data and IoT-based applications in smart environment-related fields. In this paper, a systematic review was conducted using the PRISMA approach and its selection process of identification, screening, eligibility, and inclusion, which we reported on in detail. 45 works were selected from the 240 initially extracted, based on their relevance to the six research questions we developed, and our survey of these 45 papers led us to realize the integration of big data and IoT technologies has led to great opportunities for smart environment applications (e.g. smart environment monitoring, smart metering, smart farming/agriculture, and disaster alerting) for both various end-users. However, most of these contributions are newly developed and the majority still require optimization, which can be an opportunity for researchers interested in this field.

#### References

- [1] Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. **Future Gener Comp Sy** **2013**, 29, 1645-1660.
- [2] Chin, J., Callaghan, V., Lam, I. Understanding and personalising smart city services using machine learning, the internet-of-things and big data. In Proceedings of the IEEE 26th International Symposium on Industrial Electronics (ISIE) **2017**, June, 2050-2055. URL:<https://ieeexplore.ieee.org/document/8001570>
- [3] Romero, C. D. G., Barriga, J. K. D., Molano, J. I. R. Big data meaning in the architecture of IoT for smart cities. In *International Conference on Data Mining and Big Data*; Springer, Cham **2016**, 5, 457-465.
- [4] Atzori, L., Iera, A., Morabito, G. The internet of things: A survey. *Computer networks* **2010**, 54, 2787-2805.
- [5] ITU Strategy and Policy Unit 2005. ITU Internet Reports: The Internet of Things. In Proceedings of the International Telecommunication Union (ITU), Geneva, **2012**.
- [6] Evans, D. The internet of things: How the next evolution of the internet is changing everything. *CISCO white paper* **2011**, 1, 1-11.

- [7] Yang, L., Yang, S. H., Plotnick, L. How the internet of things technology enhances emergency response operations. *Technol Forecast Soc* **2013**, 80, 1854-1867.
- [8] Khorshed, M. T., Sharma, N. A., Kumar, K., Prasad, M., Ali, A. S., Xiang, Y. (2015, December). Integrating Internet-of-Things with the power of Cloud Computing and the intelligence of Big Data analytics—A three layered approach. *APWC on CSE IEEE* **2015**, 1-8.
- [9] Shadroo, S., Rahmani, A. M. Systematic survey of big data and data mining in internet of things. *Computer Networks* **2018**, 139, 19-47.
- [10] Katal, A., Wazid, M. and Goudar, R.H., 2013, August. Big data: issues, challenges, tools and good practices. In Proceedings of the Sixth international conference on contemporary computing (IC3) *IEEE*, **2013**, 404-409.
- [11] Chebbi, I., Boulila, W., Farah, I. R. Big data: Concepts, challenges and applications. In *Computational Collective Intelligence*; Springer, Cham. 2015, 638-647.
- [12] Alam, F., Mehmood, R., Katib, I., Albogami, N. N., Albeshri, A. Data fusion and IoT for smart ubiquitous environments: a survey. *IEEE Access* **2017**, 5, 9533-9554.
- [13] Cook, D., Das, S. K. *Smart environments: technology, protocols, and applications*. John Wiley & Sons; Hoboken, New Jersey 2004, 1-10.
- [14] Cicirelli, F., Guerrieri, A., Mastroianni, C., Spezzano, G., Vinci, A. (Eds.). The Internet of Things for Smart Urban Ecosystems. *Springer*, **2019**.
- [15] Gomez, C., Chessa, S., Fleury, A., Roussos, G., Preuveneers, D. Internet of Things for enabling smart environments: A technology-centric perspective. *JAISE* **2019**, 11, 23-43.
- [16] Bibri, S. E. The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable Cities and Society* **2018**, 38, 230-253.
- [17] Zhao, S, Li, S., Da Xu, L. The internet of things: a survey. *Inform Syst Front* **2015**, 17, 243-259.
- [18] Akoka, J., Comyn-Wattiau, I., Laoufi, N. Research on Big Data—A systematic mapping study. *Computer Standards & Interfaces* **2017**, 54, 105-115.
- [19] Shah, S. A., Seker, D. Z., Hameed, S., Draheim, D. The Rising Role of Big Data Analytics and IoT in Disaster Management: Recent Advances, Taxonomy and Prospects. *IEEE Access* **2019**, 7, 54595-54614.
- [20] Talavera, J. M., Tobón, L. E., Gómez, J. A., Culman, M. A., Aranda, J. M., Parra, D. T., Garreta, L. E. Review of IoT applications in agro-industrial and environmental fields. *Comput Electron Agr* **2017**, 142, 283-297.
- [21] Al Mamun, M. A., Yuce, M. R. Sensors and Systems for Wearable Environmental Monitoring towards IOT-enabled Applications: A Review. *Sensors Journal* **2019**, 18, 7771-7788.
- [22] Fu Gu, Buqing Ma and Jianfeng Guo “Internet of things and Big Data as potential solutions to the problems in waste electrical and electronic equipment management: An exploratory study”, Proceeding of the Waste management, Pages 434–448, 2017
- [23] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA statement. *Annals of internal medicine*, 4 **2009**, 264–269.
- [24] Peffers, K., Tuunanen, T., Rothenberger, M. A., Chatterjee, S. A design science research methodology for information systems research. *JMIS* **2007**, 24(3), 45-77.
- [25] Dupont, C., Sheikhalishahi, M., Biswas, A. R., Bures, T. IoT, big data, and cloud platform for rural African needs. In Proceedings of the IST-Africa Week Conference (IST-Africa); *IEEE* **2017**, 5, 1-7.
- [26] Zhang, Y., Pu, H. Environmental indicators of sustainable computing applications for Smart City. *Concurrency and Computation: Practice and Experience* **2019**, 31(9), 4751.

- [27] Chebbi, I., Boulila, W., Farah, I. R. (2016, March). Improvement of satellite image classification: Approach based on Hadoop/MapReduce. In Proceedings of the 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP); *IEEE* **2016**,3, 31-34.
- [28] Sood, S. K., Sandhu, R., Singla, K., Chang, V. IoT, big data and HPC based smart flood management framework. *Sustain Comput-Infor* **2018**, 20, 102-117.
- [29] Yang, C., Su, G., Chen, J. Using big data to enhance crisis response and disaster resilience for a smart city. In Proceedings of the IEEE 2nd International Conference on Big Data Analysis (ICBDA) **2017** , 504-507).
- [30] Boulila, W., Farah, I. R., Hussain, A. . A novel decision support system for the interpretation of remote sensing big data. *Earth Science Informatics* **2018**, 11, 31-45.
- [31] Rajeswari, S., Suthendran, K., Rajakumar, K. A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics. In Proceedings of the 2017 International Conference on Intelligent Computing and Control (I2C2) ; *IEEE* **2017**,6, 1-5.
- [32] Zeinab, K. A. M., Elmustafa, S. A. A. Internet of Things applications, challenges and related future technologies. *World Scientific News* **2017**, 2, 126-148.
- [33] Narayanan, K. Addressing the challenges facing IoT adoption. *Microw. J* **2017**, 60, 110-118.
- [34] Banafa, A. (2017). Three major challenges facing iot. *IEEE IoT Newsletter* **2017**.
- [35] D. Ardagna, C. Cappiello, W. Sam, M. Vitali, Context-aware data quality assessment for big data. *Future Gener Comput Sys* **2018**, 89 548 – 562.
- [36] Boulila, W. A top-down approach for semantic segmentation of big remote sensing images. *Earth Science Informatics* **2019**, 1-12.
- [37] Xu, B., Wang, W., Wu, Y., Shi, Y., Lu, C. Internet of things and big data analytics for smart oil field malfunction diagnosis. In Proceedings of the IEEE 2nd International Conference on Big Data Analysis (ICBDA) **2017**, 6,178-181.
- [38] Huang, Y., Zhao, Q., Zhou, Q., Jiang, W. Air quality forecast monitoring and its impact on brain health based on big data and the Internet of Things. *IEEE Access* **2018**, 6, 78678-78688.
- [39] Vuran, M. C., Salam, A., Wong, R., Irmak, S. Internet of underground things in precision agriculture: Architecture and technology aspects. *Ad Hoc Networks* **2018**, 81, 160-173.
- [40] Lieberman, J., Leidner, A., Percivall, G., Rösndorf, C. (2017, December). Using big data analytics and IoT principles to keep an eye on underground infrastructure. In Proceedings of the IEEE International Conference on Big Data (Big Data) **2017**, 4592-4601.
- [41] Babar, M., Arif, F. Smart urban planning using Big Data analytics to contend with the interoperability in Internet of Things. *Future Gener Comp Sy* **2017**, 77, 65-76.
- [42] Lu, S. Q., Xie, G., Chen, Z., Han, X. The management of application of big data in internet of thing in environmental protection in China. In Proceedings of the IEEE First International Conference on Big Data Computing Service and Applications **2015**, 3, 218-222.
- [43] Fang, S., Da Xu, L., Zhu, Y., Ahati, J., Pei, H., Yan, J., Liu, Z. An integrated system for regional environmental monitoring and management based on internet of things. *IEEE T Ind Inform* **2014**, 10, 1596-1605.
- [44] Baranwal, T., Pateriya, P. K. Development of IoT based smart security and monitoring devices for agriculture. In Proceedings of the IEEE International Conference-Cloud System and Big Data Engineering (Confluence) **2016**, 597-602.
- [45] Din, S., Ghayvat, H., Paul, A., Ahmad, A., Rathore, M. M., Shafi, I. (2015, December). An architecture to analyze big data in the internet of things. In Proceedings of the 9th International Conference on Sensing Technology (ICST); *IEEE* **2015**, 677-682.
- [46] <https://flume.apache.org>

- [47] <https://kafka.apache.org>
- [48] Villari, M., Celesti, A., Fazio, M., Puliafito, A. AllJoyn Lambda: An architecture for the management of smart environments in IoT. In 2014 International Conference on Smart Computing Workshops; IEEE **2014**, 11, 9-14.
- [49] Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D.B., Amde, M., Owen, S. and Xin, D. Mllib: Machine learning in apache spark. *JMLR* **2016**, 17, pp.1235-1241.
- [50] Berouine, A., Lachhab, F., Malek, Y. N., Bakhouya, M., Ouladsine, R. . A smart metering platform using big data and IoT technologies. In Proceedings of the 3rd international conference of cloud computing technologies and applications (CloudTech);IEEE **2017**, 10, pp. 1-6.
- [51] Thorat, A., Kumari, S., Valakunde, N. D. An IoT based smart solution for leaf disease detection. In Proceedings of the IEEE International Conference on Big Data, IoT and Data Science (BIGDATA) **2017**, 193-198.
- [52] Fazio, M., Celesti, A., Puliafito, A., Villari, M. Big data storage in the cloud for smart environment monitoring. *Procedia Comput. Sci* **2015**, 52, 500-506.
- [53] Corbellini, A., Mateos, C., Zunino, A., Godoy, D., Schiaffino, S. Persisting big-data: The NoSQL landscape. *Information Systems* **2017**, 63, 1-23.
- [54] Kang, Y. S., Park, I. H., Rhee, J., Lee, Y. H. MongoDB-based repository design for IoT-generated RFID/sensor big data. *IEEE Sensors Journal* **2015**, 16, 485-497.
- [55] Suci, G., Anwar, M., Ganaside, A., Scheianu, A. IoT time critical applications for environmental early warning. In Proceedings of the IEEE International Conference on Electronics, Computers and Artificial Intelligence (ECAI) **2017**, 6, 1-4
- [56] Elarabi, T., Deep, V., Rai, C. K. (2015, December). Design and simulation of state-of-art ZigBee transmitter for IoT wireless devices. In Proceedings of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) **2015**, 12, 297-300.
- [57] Pan, G., He, J., Wu, Q., Fang, R., Cao, J., Liao, D. Automatic stabilization of Zigbee network. In Proceedings of the IEEE International Conference on Artificial Intelligence and Big Data (ICAIBD) **2018**, 6, 224-227.
- [58] Alshamsi, Anwar, et al. "Monitoring pollution: Applying IoT to create a smart environment." proceeding of the 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA). IEEE, 2017.
- [59] Lavric, A., Petrariu, A. I. LoRaWAN communication protocol: The new era of IoT. In Proceedings of the IEEE International Conference on Development and Application Systems (DAS) **2018**, 5, 74-77.
- [60] A. J. Wixted, P. Kinnaird, H. Larijani, A. Tait, A. Ahmadinia, and N. Strachan. Evaluation of LoRa and LoRaWAN for wireless sensor networks. *IEEE SENSORS* **2016**, 1-3.
- [61] Carvalho, D. F., Depari, A., Ferrari, P., Flammini, A., Rinaldi, S., Sisinni, E. On the feasibility of mobile sensing and tracking applications based on lpwan. In Proceedings of the IEEE Sensors Applications Symposium (SAS) **2018**, 3, 1-6
- [62] Warnakulasooriya, K., Jayasuriya, Y. P., Sudantha, B. H. Generic IoT Framework for Environmental Sensing Researches: Portable IoT Enabled Weather Station. In Proceedings of the IEEE International Conference on System Science and Engineering (ICSSE) **2018**, 1-5.
- [63] Chaczko, Z., Kale, A., Santana-Rodríguez, J. J., Suárez-Araujo, C. P. (2018, June). Towards an IOT Based System for Detection and Monitoring of Microplastics in Aquatic Environments. In Proceedings of the IEEE 22nd International Conference on Intelligent Engineering Systems (INES), **2018**, 57-62.
- [64] Srinivasulu, P., Babu, M. S., Venkat, R., Rajesh, K. (2017, April). Cloud service oriented architecture (CSOA) for agriculture through internet of things (IoT) and big data. In Proceedings of the IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE) **2017**, 4, 1-6.
- [65] Roy, S., Ray, R., Roy, A., Sinha, S., Mukherjee, G., Pyne, S., Hazra, S. (2017, August). IoT, big data science & analytics, cloud computing and mobile

- app based hybrid system for smart agriculture. In Proceedings of the IEEE Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON) **2017**, 303-304.
- [66] Koo, D., Piratla, K., Matthews, C. J. Towards sustainable water supply: schematic development of big data collection using internet of things (IoT). *Procedia engineering* **2015**, 118, 489-497.
- [67] H Wang, H., Osen, O. L., Li, G., Li, W., Dai, H. N., Zeng, W. Big data and industrial internet of things for the maritime industry in northwestern norway. In TENCON 2015-2015 IEEE Region 10 Conference **2015**, 1-5.
- [68] Babar, M., Arif, F., Jan, M. A., Tan, Z., Khan, F. Urban data management system: Towards Big Data analytics for Internet of Things based smart urban environment using customized Hadoop. *Future Gener Comp Sy* **2019**, 96, 398-409.
- [69] Chen, M., Yang, J., Hu, L., Hossain, M. S., Muhammad, G. Urban healthcare big data system based on crowdsourced and cloud-based air quality indicators. *IEEE Communications Magazine* **2018**, 56, 14-20.
- [70] Xu, B., Wang, W., Wu, Y., Shi, Y., Lu, C. Internet of things and big data analytics for smart oil field malfunction diagnosis. In Proceedings of the IEEE 2nd International Conference on Big Data Analysis (ICBDA) **2017**, 6, 178-181. IEEE.
- [71] Chang, H. Y., Wang, J. J., Lin, C. Y., Chen, C. H. An Agricultural Data Gathering Platform Based on Internet of Things and Big Data. In Proceedings of the International Symposium on Computer, Consumer and Control (IS3C) **2018**, 12, 302-305.
- [72] Alam, F., Mehmood, R., Katib, I., Albogami, N. N., Albeshri, A. Data fusion and IoT for smart ubiquitous environments: a survey. *IEEE Access* **2017**, 5, 9533-9554.
- [73] Lazarescu, M. T. Design of a WSN platform for long-term environmental monitoring for IoT applications. In Proceedings of the IEEE Journal on emerging and selected topics in circuits and systems **2013**, 3(1), 45-54.
- [74] Kadir, E. A., Efendi, A., Rosa, S. L. Application of LoRa WAN Sensor and IoT for Environmental Monitoring in Riau Province Indonesia. *Proceeding of the Electrical Engineering Computer Science and Informatics* **2018**, 5(5), 281-285.
- [75] Lee, J. G., Tsai, W. F., Lee, L. C., Lin, C. Y., Lin, H. C., Tsuang, B. J. In-place query driven big data platform: Applications to post processing of environmental monitoring. *Concurrency and Computation: Practice and Experience* **2017**, 29(13), 4135.
- [76] Onal, A. C., Sezer, O. B., Ozbayoglu, M., Dogdu, E. Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning. In Proceedings of the IEEE International Conference on Big Data *2017*, 12, 2037-2046.
- [77] He, Z., Yang, F., Li, Z., Liu, K., Xiong, N. Mining Channel Water Depth Information From IoT-Based Big Automated Identification System Data for Safe Waterway Navigation. *IEEE Access* **2018**, 6, 75598-75608.
- [78] Zhang, F., Xue, H. F., Zhang, J. C. Multi-source big data dynamic compressive sensing and optimization method for water resources based on IoT. *Sustainable Computing: Informatics and Systems* **2018**, 20, 210-219.
- [79] Middel, A., Lukasczyk, J., Zakrzewski, S., Arnold, M. and Maciejewski, R. Urban form and composition of street canyons: A human-centric big data and deep learning approach. *Landscape and urban planning* **2019**, 183, 122-132.
- [80] Al-Sarem, M., Boulila, W., Al-Harby, M., Qadir, J. and Alsaeedi, A. Deep Learning-Based Rumor Detection on Microblogging Platforms: A Systematic Review. *IEEE Access* **2019**, 7, 152788-152812.
- [81] Chin-Feng Tsai, Ting-Wen Liang, "Application of IoT Technology in The Simple Micro-farming Environmental Monitoring", proceeding of the 2018 IEEE International Conference on Advanced Manufacturing (ICAM), pages: 170-172, 2018.
- [82] Pallavi, S., Jayashree D. Mallapur, and Kirankumar Y. Bendigeri. "Remote sensing and controlling of greenhouse agriculture parameters based on IoT."

proceeding of the 2017 International Conference on Big Data, IoT and Data Science (BIGDATA). IEEE, 2017.