REGULAR PAPER

An overview of data reduction solutions at the edge of IoT systems: a systematic mapping of the literature

Laércio Pioli[1] · **Carina F. Dorneles[1]** · **Douglas D. J. de Macedo[1]** · **Mario A. R. Dantas[2]**

Abstract

Internet of Things (IoT) is a technology that connects devices of different types and characteristics through a network. The massive quantity of the heterogeneous generated data by the sensors imposes many challenges in making these data available to IoT applications. Data reduction and preprocessing are promising concepts that help to handle these data efficiently before storing them. Applying data reduction methods at the edge has emerged as an efficient solution. In such context, this systematic mapping is intended to investigate the data reduction solutions performed exclusively at the edge through a set of research questions. To reach this objective, we performed a Systematic Literature Mapping (SLM) in which 35 papers were strictly analyzed among a total of 853 articles. Finally, we present the results of these analyses answering questions that relate to the researcher's used techniques, hardware technologies, used data type, and contributed objects to perform the data reduction techniques on the edge of the IoT systems.

Keywords IoT · Data reduction · Edge · Fog · Big data · Systematic mapping

Mathematics Subject Classification 68-02 · 68T09 · 68U01

✉ Laércio Pioli
laercio.pioli@posgrad.ufsc.br

Carina F. Dorneles
carina.dorneles@ufsc.br

Douglas D. J. de Macedo
douglas.macedo@ufsc.br

Mario A. R. Dantas
mario.dantas@ice.ufjf.br

[1] Computer Science, Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina 88040-900, Brazil

[2] Computer Science, Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, Minas Gerais 36036-900, Brazil

1 Introduction

Internet of Things (IoT) is a structural technology that connects plenty of devices from different types and characteristics using interconnected networks. It was initially described by [1] in 1999 when introducing device-to-device communication at Procter & Gamble. According to [2], IoT has a combination of three elements: things, internet, and semantic. It allows several sensors to communicate in the most varied possible domains, thus presenting a considerable potential to generate insights over the data that travels through the network. Thus, receiving much attention in the last decade from the academy and industry. Moreover, the idea of connecting things through a network domain contributed to the data growth. With numerous sensors and scattered data sources worldwide, the amount of generated, processed, and stored data has increased considerably. However, this ever-increasing data generation imposes some challenges in the process of providing and consuming these data. In this case, data preprocessing and reduction are promising concepts that help to manage this data efficiently, increasing resource consumption and saving money.

To support such activities, the concept of edge computing, which handles data closer to the generated point before being sent to the cloud, has emerged as an efficient solution. Although edge computing principles can be observed in the 60s, in the 1990s the authors [3] presented their network solution due to the web congestion problem. They spread 12,000 servers around 1,000 networks to solve such an issue. This topology improves data effectiveness, reducing bottlenecks on the cloud by dealing with raw data closer to the generation point. In addition, it avoids unnecessary data being transmitted along the network, releasing it up for cleaning and consistent data. Usually, this component is placed between the raw data and the cloud servers through intelligent hardware technologies. Gateways, servers, routers, switches are some of these hardware technologies used as edge hardware mediators. However, it is possible to use any hardware with minimal processing and storage to perform this task. These gateways that connect devices with a wide variety of protocols to the cloud fulfill their data management by increasing the computation and reducing their storage.

The concept of applying data reduction at the edge may bring many advantages when dealing with IoT sensor data. Bandwidth network latency can be reduced in a gateway, thus preventing I/O bottlenecks on the overall network connectivity. Many data reduction techniques can be used in the three-phase stages. Examples of these techniques can be found in Sect. 4. In [4] the authors presented the three major on-the-fly data reduction stages. The first one is pattern detection and removal. The second stage is the inline deduplication, and the third one is the inline compression function, both implemented in the system node. These stages occur following the logic flow presented in Fig. 1. After the user writes the command, the generation of a computed hash to this write is performed. Next, the pattern matching detection and removal stage begins. If the founded pattern is known, the process generates a hash ID to this data and proceeds to the next user's writing ID. If this pattern is not known, a hash search deduplication should be created. The process of finding a match is then started. When the match occurs, the relationship is established, and metadata is written. When it does not occur, the hash is introduced to the database, the data is then compressed, and the data is written. Figure 1 summarizes these steps. Data reduction at the edge has the
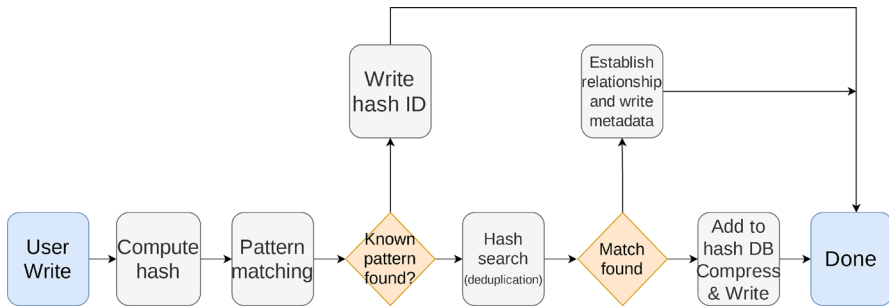
Fig. 1 Data reduction logic process [4]

potential to reduce bottlenecks of an IoT system through four different perspectives [5]. They are network bandwidth, network energy, I/O throughput, cloud storage, and traffic costs.

The increase in data production by different classes of heterogeneous IoT sensors is imposing many challenges regarding data management. Distinct classes of technologies and domains that incorporate the IoT paradigm (e.g., smart home, smart cities, smart grids, mobile applications, smart devices, machine-generated data, wearable devices, and applications, etc.) collaborate to this data deluge producing data in a variety of ways. In such a context, the problem of forwarding yottabytes of heterogeneous IoT sensor data directly to the cloud without any data reduction technique is addressed in this study. Applying data reduction treatments avoids network and systems bottlenecks, feeding the up-layered applications with relevant and useful data. This process transforms fuzzy data by many approaches and techniques into a corrected, ordered, and simplified form, reducing invalid data and preparing it for different types of applications. The objective of this research is to present the principal and actual solutions regarding data reduction at the edge of IoT systems through a systematic mapping of the literature. We highlight the used data reduction techniques, the different data formats, the main researcher's proposals, the architectural implementation choice, and finally, the hardware technologies they used to perform data reduction.

In this article, we presented a systematic mapping of the literature in which we investigated how the data reduction methods are being applied at the edge of IoT systems. Our main contributions are: (i) A data reduction surveys comparison table, in addition, the reviewing of existing literature works that proposes solutions regarding data reduction in different contexts, (ii) the answering of six research questions that address pertinent information on how these data reduction methods were introduced. Some of these issues deal with elements such as (a) the growth development of this area, (b) the used techniques that authors were applying to reduce the data volume at the edge, (c) the used data type for experimentation, (d) the proposed contributions by the authors, (e) analyses about the number of components of the author's proposed solutions, (f) the used technologies to perform authors data reduction solutions. (iii) the discussion of elemental open issues regarding data reduction.

The remainder of this paper is structured as follows: Section 2 presents some related works concerning the importance of data reduction in different contexts, in addition,

to a data reduction surveys comparison table. Section 3 highlights the research method used in this work. Section 3.1 presents the mapping planning used to perform the SLM here presented. The PICOC guideline is presented in Sect. 3.2. Further, the research questions and the search strategy are presented in Sects. 3.3 and 3.4 respectively. Section 3.5 presents the data extraction process. Moreover, the systematic mapping report in which we present our results and interesting discussions is presented in Sect. 4. Threats to the validity of this research followed by interesting open issues are presented in Sects. 5 and 6 respectively. Finally, Section 7 conclude this study.

2 The importance of the data reduction

Although we are looking for improvement solutions that directly attack the data reduction technique performed at the edge layer of the IoT technology, it is important to present the state-of-the-art and development trends of it. First, in this section, we present an interesting survey comparison in Table 1 highlighting some observations of the researches when dealing with the data reduction approach. Next, we present and discuss research studies that employ data reduction in specific contexts as part of its solution. In such a case, these studies are not focused directly on improving the data reduction technique field, instead, they focus on different contextualized research fields areas (e.g., data analysis, data management, network traffic reduction, power consumption, etc). However, the importance of presenting such studies enforces the need to keep improving data reduction techniques, which are used as a preprocessing tool in a wide variety of domains.

Our systematic mapping of the literature is needed because it condenses the researchers that implement the data reduction solutions at the edge layer of IoT systems as the first goal of the study. To the best of our knowledge, there is no research in this specific area that evaluates data reduction methods in IoT systems considering the edge layer. We present now the majority of observations about interesting research survey studies found in the literature, which consider perspectives regarding the data reduction technique. These published surveys are presented in Table 1. An interesting observation regarding all studies, except ours, is that they do not were built considering an auditable and consolidated systematic guideline. The usage of this approach contributes to the quality of the study, proposing to the readers the possibility to verify and check the presented results.

On one hand, some authors discussed the usage of data reduction in the context of Big data paradigm. Researchers in [6] reviewed the complexity of big data and the need to implement data reduction methods in such context. They also presented a taxonomy of the big data reduction methods including dimension reduction, data compression, Machine Learning (ML), and data mining. Authors in [7] discussed the preprocessing data workflow also in the context of big data. They presented discussions regarding the four phases of data preprocessing (e.g., cleansing, integration, reduction, and transformation)

On the other hand, some surveys focus on the Wireless Sensor Networks (WSN) technology. The work [12] reviewed the specific data aggregation technique in IoT sensor networks. More specifically, they presented many data aggregation techniques

Table 1 Data reduction surveys comparison

Work	Year	Systematic Search	General Techniques Review	Focused on the Edge	IoT Context
[6]	2016	✗	✓	✗	✗
[7]	2017	✗	✓	✗	✗
[8]	2018	✗	✓	✓	✓
[9]	2019	✗	✗	✗	✓
[10]	2019	✗	✗	✗	✓
[11]	2020	✗	✗	✓	✓
[12]	2020	✗	✗	✗	✓
[13]	2021	✗	✗	✗	✓
[14]	2021	✗	✓	✗	✓
Our	2021	✓	✓	✓	✓

and protocols of the WSN IoT category. Authors in [13] discussed the different data aggregation methods and protocols implemented in WSN in a non-systematic survey methodology.

Finally, all the other authors considered the IoT technology to present their survey. In [9] the specific technique of anomaly detection in the IoT context is discussed. They discussed the challenges when applying such a technique to IoT stream data. Authors in [8] refer to the Industrial Internet of Things (IIoT). This category has different and distinct properties and requirements compared to the traditional IoT. In such context, it is focused on be benefices of the IoT to improve productivity, efficiency, safety, and industrial intelligence. Researchers in [10] discussed the data fusion techniques in the IoT context without using a systematic methodology to discuss the studies, though. The focus was on the different IoT application domains, presenting special attention to security and privacy. The authors [11] focused on the usage of edge computing to support smart city development. They review the state-of-the-art literature of the edge computing application in smart cities, proposing a taxonomy to classify such context. The study [14] focused on IoT data preprocessing, discussing the four data preprocessing techniques, which includes data reduction. In our case, a SLM focused on data reduction solutions where the reduction concept is the first goal of the studies and is applied specifically at the edge layer of IoT systems.

Different studies areas tackle the data reduction technique. WSN and Artificial Intelligence (AI) are two of them that received our attention due to their interesting growing projection.

On one hand, authors proposed solutions to reduce the generated data in the context of WSN from different perspectives [15,16], and [17]. In [16] a data reduction algorithm at the sensor-node level that adapts the frame rate and reduces the number of images sent from the sensor node to the coordinator was proposed. The study [15] presented a two-tier data reduction framework that implements a gradient-based model. In the framework, a dual prediction (DP) scheme is implemented at both the
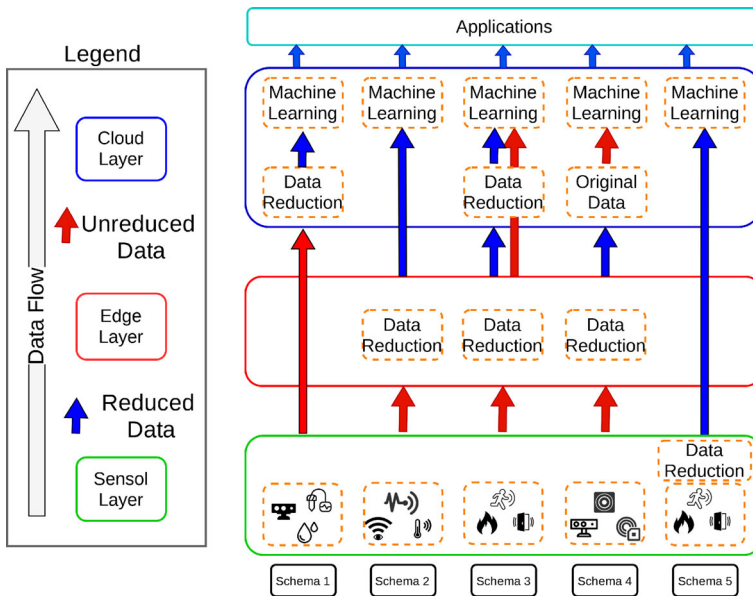
Fig. 2 Edge-cloud architectural schema

sensor nodes and the cluster head (CH). In such a case, the data compression (DC) scheme is applied between the CHs and the sink node.

In the context of AI there are some interesting studies that make use of the data reduction technique to improve the data quality [18–20], and [21]. The study [19] presented a solution that merges cloud and edge computing, aiming to solve the architectural problem of data analytics. They presented a deep learning-based approach to perform data reduction on the edge before performing the machine learning techniques on the cloud. As mentioned by the authors, the data reduction functionality is located on the edge to implement the data reduction and then sent to the cloud where the data are deeply explored using the decoder part of the autoencoder. Researchers in [21] proposed a framework that uses autoencoders model at the edge to apply data compression on the raw data, reducing then the high-dimensional data into compact representation. The majority of the presented works might fall into one of the architectural schema presented in Fig. 2.

Implementing data reduction in different edge-cloud architectural layers to further perform data analytic techniques in the cloud has proved to be a consolidated approach. However, there are also other possibilities when distributing these techniques on the architectural layer.

Sending the generated sensor data directly to the cloud (Schema 1) seems to be the usual idea when the objective is data processing and analytic. It happens due to the computational processing power difference of cloud servers and edge nodes. However, depending on the quantity of transmitted data, the latency might increase when sending the data directly from the sensor layer to the cloud. Therefore, an additional processing layer (i.e., edge layer) is needed.
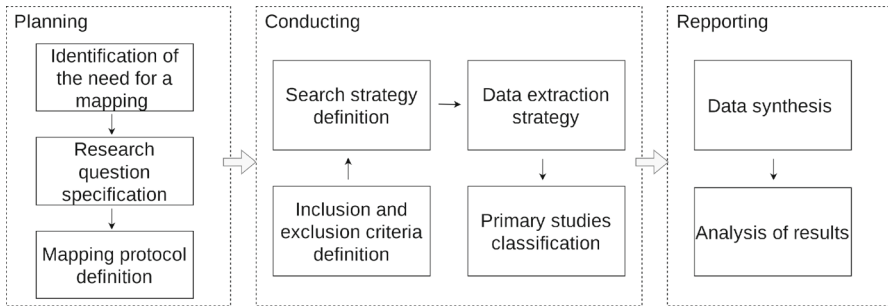
Fig. 3 Research methodology process

Preprocessing the raw data at the edge before reaching the cloud, transforming high-dimensional data into compact data (Schema 2) is an interesting solution. After reduce the data on the edge layer, they could be directly used by AI solutions in the cloud. In some cases, for ML approaches, the encoder part of the network maps input is performed at the edge layer and the decoder part is performed in the cloud. This schema reduces the network I/O bottleneck, network energy consumption, and cloud storage usage for instance.

There are also cases where the used buffer at the edge is fully used. When it happens, the raw unreduced data should be forwarded to the cloud to avoid data loss. In the cloud, it is also possible to apply a data reduction method on the data or even forward it to the ML module. This solution is called the Edge+Cloud hybrid approach and is represented by (Schema 3).

Depending on the architectural implementation and context, it is also possible to decompress the data in the cloud to get the identical generated sensor data (Schema 4). It probably would be used to decrease the network bottleneck at the edge-cloud interconnection, being an interesting solution when a massive data volume is produced and a possible bottleneck is expected in the network.

There are also solutions, [22,23] and [24] that implement data reduction directly in the sensor layer. Although these studies do not implement AI techniques in their architectures, we understand that introducing an AI layer in these solutions is possible and promissory. This architectural approach is represented by (Schema 5).

3 Systematic literature mapping (SLM) methodology

SLM is a type of secondary study that aims to aggregate empirical evidence through a methodological approach. It targets to answer questions related to a research field following a systematic methodology that can be reproduced. Intending to discover unknown results, this systematic approach tries to identify and interpret all primary researches under the studied scope. The author [25] discuss the importance of mapping studies in software engineering. This research was based on the presented SLM definition and the Fig. 3 reflects the used methodology process.

3.1 Mapping planning

Several activities should be considered before performing a review. The first activity of the planning stage relates to identifying the need for performing a systematic mapping. The science development of a specific area is improved continually by researchers, and they are constantly remodeling such fields. Due to this, the understanding area might be requested by new researchers or non-experts. Aggregating this information is a common researcher requirement because it allows analysis so that the preliminary results individually do not present such information.

Specifying the research question is another critical activity. Every research aims at solving or improving some specific problem. Essentially, the research development must be clear and concise, which helps avoid research bias introduction. In [26] researchers argue that the most crucial part of systematic mapping conduction is the research question specification process. The main objectives in the research question specification process are to identify the primary studies that will be analyzed and the extracted data that will be used to answer these questions requirements. Finally, the development of the mapping protocol, which can be understood as the act of putting the protocol into practice, was presented.

3.2 PICOC

Research questions should bounder the research scope. The researchers in [27] suggested considering an extended medical guideline, PICOC (Population, Intervention, Comparison, Outcome, Context), to construct the research questions. In this guideline, Population refers to the group of elements that are being investigated; Intervention refers to the element that addresses the study; the Comparison is used to compare the previously specified intervention; the Outcome describes the expected results; and finally, the Context element delimits the domain of the study in which the intervention is delivered and investigated. For our study, the definitions of the PICOC terms are presented.

- Population (P): Fog OR Edge Computing OR Gateway.
- Intervention (I): Data reduction.
- Comparison (C): Null
- Result (O): Solution OR Method OR Approach OR Framework OR Schema.
- Context (C): - Internet of Things.

We are particularly interested in investigating data reduction solutions that are employed at the edge of IoT systems. As we are not considering any other interventions to compare, thus, the comparison is null. As a result, we are interested in getting all the possible outcomes. Solutions, methods, approaches, frameworks, and schema are some expected results that were specified. We are considering the Fog & Edge computing context, which is where the intervention is being delivered. It is essential to mention here that we are not focusing on cloud solutions unless it has essential elements that were constructed at the edge.

3.3 Research questions

The previous PICOC guideline is a component that ought to be built according to the interest of the referred research. In such a case, we are interested in analyzing how data reduction, which is the intervention of our guideline, is being addressed regarding at the edge of the internet of things domain. There, we are looking for outcomes such as solutions, methods, approaches, frameworks, or schema that aim to reduce the data quantity specified at the edge context. The previously selected keywords (defined for the PICOC construction) allow us to create and specify the research question we are interested in solving. It enables us to explore the bounded questions here presented. After defining the research questions, we can identify the primary studies selected to be analyzed. The authors in [28] advocate that research questions might be clear and narrow. Below, we present the set of Mapping Questions (MQ) that delimit the scope of this mapping research.

- MQ1: How many studies have been published over the years, and how this distribution was constructed?
- MQ2: Which techniques researchers are applying to perform data reduction at the edge?
- MQ3: Which data type researchers are using for experimentation?
- MQ4: Which was the contributed object proposed by the researchers?
- MQ5: Where are the authors implementing the data reduction solutions?
- MQ6: Which hardware technologies were used by authors' to perform their data reduction solutions?

3.3.1 Inclusion/Exclusion criteria

The objective of the inclusion and exclusion criteria is to select the research studies that fit the investigated questions. We present the inclusion criteria used to find the primary studies. The exclusion criteria are the negative form of the defined inclusion criteria, and because of that, they were omitted.

Inclusion Criteria:

IC1: Articles that propose a data reduction solution at the edge layer considering the IoT technology AND
IC2: Articles that necessarily have a title and abstract AND
IC3: Articles published in English AND
IC4: Articles published at peer-review events, such as workshops OR conferences OR journals AND
IC5: Articles published after 2015.

3.4 Search strategy

In this section, we defined the search strategy elements used in this research. The presented libraries and the PICOC guideline methodology shown in Sect. 3.2 were used in this process.

Table 2 Used search string on the research

Search String
("Fog" OR "Edge" OR "Gateway") AND
("Data Reduction") AND
("solution" OR "method" OR "approach" OR "framework" OR "schema") AND
("IoT" OR "Internet of Things")

3.4.1 Sources

We specify the source libraries used to perform the search after defining the PICOC elements, the research questions, and the inclusion/exclusion criteria. ACM Digital Library, EI Compendex, IEEExplore, ScienceDirect, Scopus, and Springer were the six selected libraries that were defined using the following criteria: (i) the ability to perform searching using logical characters (e.g., "AND" and "OR"); (ii) to allow searching in the body of the document and metadata fields (e.g., title, keyword, and abstract); and (iii) to be repositories having computer science and engineering contents.

3.4.2 String search

Aiming to find the articles that propose the data reduction solutions applied in the context of the internet of things technology, the used string search is now explored in Table 2. The first set of elements of the string search was directly related to the population presented on PICOC, which is composed of three elements. Edge computing and Fog computing are architectural models usually used to reduce data quantity through data reduction techniques in IoT technology. For this reason, we selected these terms in addition to the synonym "Gateway" that compose the population of our defined PICOC guideline. The second line refers to the element that addresses the study and refers to the intervention presented in Sect. 3.2. As in this paper, we are investigating "Data Reduction" solutions applied in IoT systems and architectures, the second line of the table presents this single term. Moreover, the third line refers to the possible outcomes in which the intervention can be delivered. We specified similar synonyms for the outcome, advocating that these terms cover more of the different described solutions. Finally, the last line of the string refers to the defined context we are investigating. A similar synonym (e.g., "IoT") has also been added to the original "Internet of Things" term, thus aiming at greater effectiveness in returning primary searches. It increases the quality of the returned studies, thus reducing the quantity of the outer-bound paper.
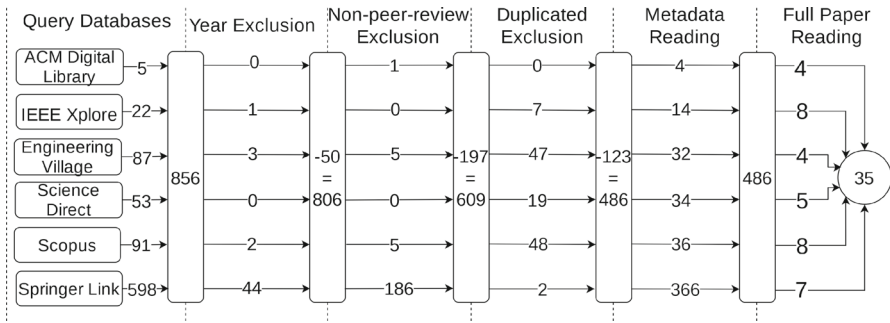
Query Databases	Year Exclusion	Non-peer-review Exclusion	Duplicated Exclusion	Metadata Reading	Full Paper Reading
ACM Digital Library —5→	0 →	1 →	0 →	4	4
IEEE Xplore —22→	1 →	0 →	7 →	14	8
Engineering Village —87→	3 → -50 =	5 → -197 =	47 → -123 =	32	4
	856	806	609	486	35
Science Direct —53→	0 →	0 →	19 →	34	5
Scopus —91→	2 →	5 →	48 →	36	8
Springer Link —598→	44 →	186 →	2 →	366	7

Fig. 4 Data extraction process

3.5 Data extraction process

The data extraction process defined in this survey is composed of six main steps and shows that 35 primary studies were accepted for detailed analysis. The main steps are presented in Fig. 4. We summarized all these studies in a table available on this link [1].

First, stage one represents the number of returned primary studies after executing the string in the libraries presented in Sect. 3.4.1. Although we configured all the libraries to perform the search string considering only the metadata fields (i.e. title, keywords, and abstract), the ACM library has allowed us to perform the search string only on the abstract field. Next, in the second stage, we have selected the papers published after 2015 constituting a full five-year publishing time range. Five years of time range provide us the most recent results while removing the papers that are not part of the scope. Then, in stage three, we removed 197 files due to the EC5 criteria, which removes non-peer-reviewed channels papers (e.g., Books Chapters, Book Reviews, Editorials, Mini review, Info, Other, etc.). After that, in stage four, we have removed all the 123 duplicated articles because simultaneous libraries can index them. Then, in stage five, we have read all the papers' metadata (i.e., keywords, titles, and abstracts) of the overall quantity of selected papers. Finally, in stage six, we take a more rigorous reading, finding 35 papers that covered the scope of our study.

4 Systematic literature mapping report

After selecting the bounded papers, the detailed results of the SLM can be presented. The 35 primary studies were deeply analyzed and classified to answer the mapping questions presented in Sect. 3.3.

---

[1] https://docs.google.com/spreadsheets/d/1A21KEEeS9_SEmERRJ6aKOr6mC-khA7v1e0ARbyTnyYg/edit?usp=sharing.
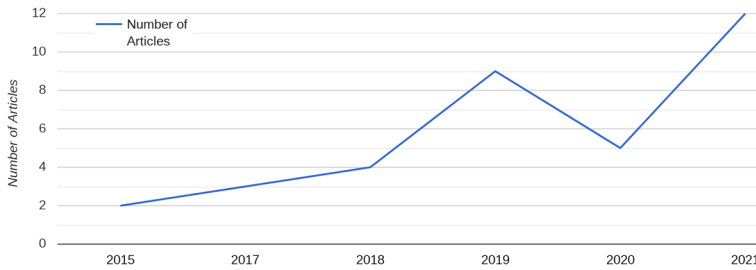
Fig. 5 Accepted papers years distribution

4.1 Studies distribution

Figure 5 addresses the question *(MQ1: How many studies have been published over the years, and how this distribution was constructed?)*, plotting the number of papers published in each year.

It is possible to see in Fig. 5 a slight growth in solutions that propose a data reduction at the edge environments since 2015. The growing demand for big data solutions contributes to the usage and generation of data reduction approaches in systems architecture. The data preprocessing performed on the edge layer is fundamental to reducing systems bottleneck, improving the overall applications' performance. Analyzing the graph, we note that 2015 was the year with the lowest number of published studies, thus forming a lower limit for publications. After 2015, the graph trend has increased for all years, excepting 2020. More specifically, in 2017 we find 3 studies, in 2018 we find 4 studies, followed by 9 in 2019. However, in 2020 we faced a considerable decrease in the proposed solutions. We suggest that this decrease in 2020 is a reflection of the COVID-19 pandemic disease the entire world faced at the end of 2019. The authors, [29], presented a study discussing how the global COVID-19 pandemic has affected the world, including the IoT technology. At the end of 2021, 12 more studies were published. This indicates an increase in the research area, justifying the need for this study that can help solve emerging problems.

4.2 Data reduction techniques

Figure 6 addresses the question *(MQ2: Which techniques researchers are applying to perform data reduction at the edge?)*. Throughout this research question, we can understand, in general, which are the most used techniques to perform a data reduction at the edge. These results bring us interesting information once the hardware used at the edge has less processing power than the used in the cloud. For instance, data filtering 18.8% and data compression 16.7% were the most used techniques to reduce data volume at the edge.

The data filtering technique that was the first most used data reduction technique is proposed in [30]. The authors presented a filtering solution employing an SSN ontology and its instances. The data reduction is performed in the gateway through a structure to "prune" the non-contextual data. received sensor data are then filtered
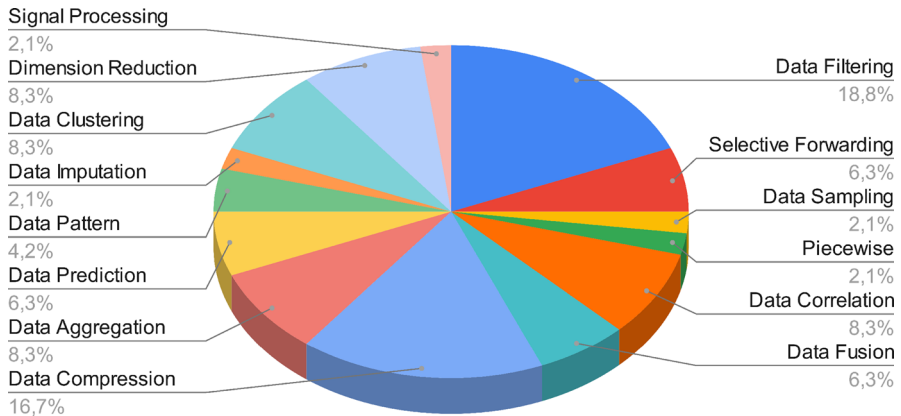
Fig. 6 Used technique

and sent to the cloud. Some solutions employ algorithms to filter data. The work [31] presented a concept and an algorithm to use redundant IoT-gateway hardware at the edge layer to increase availability and fault tolerance of data reduction by a factor of two. The proposed filtering algorithm implemented a cooperative approach to adapt to this concept, using PIP to perform data reduction. Authors in [32] proposed a multi-tier data reduction solution. In the gateway tier, the authors implemented a data filtering algorithm based on the PIP technique and included additional features, namely, interval restriction, dynamic caching, and weighted selection. At the edge, they implemented a data fusion algorithm that is based on the optimal set. In this step, the authors used the mean squared error (MSE) and recovery accuracy as the evaluation index. The work [33] proposed a data filtering and fusion in-networking approach to reduce the IoT sensor data. The data filtering solution is based on data change detection and deviation and compose the first step of the solution. After filtering, in the second layer, it occurs the data fusion is based on a minimum square error.

The data compression technique was used from different perspectives. Solutions that involve compression algorithms were also presented. The authors [34] proposed the implementation of some algorithms to reduce data transmission time among the network. They used the RAKE data compression algorithm to reduce memory when the new data value is different from the one in the gateway stored. In [24] the researchers presented a two-tier data reduction approach with different algorithms solutions. At the first sensor nodes tier, the authors apply data compression algorithms, and in the second tier network gateways, these gateways performed a clustering algorithm on the sensor nodes data sets. The authors [35] presented a parallel implementation of a History Principal Component Analysis (PCA) algorithm aiming at data compression. They evaluated their solution on a real-life Structural Health Monitoring (SHM) system to provide information about building conditions.

Architectural and framework solution that involves compression methods was also noticed. The researchers [36] proposed an architecture for transmitting fewer data and using low storage. The authors' solutions distributed multiple industrial computers in different factory areas as edge devices for data collection and compression. In
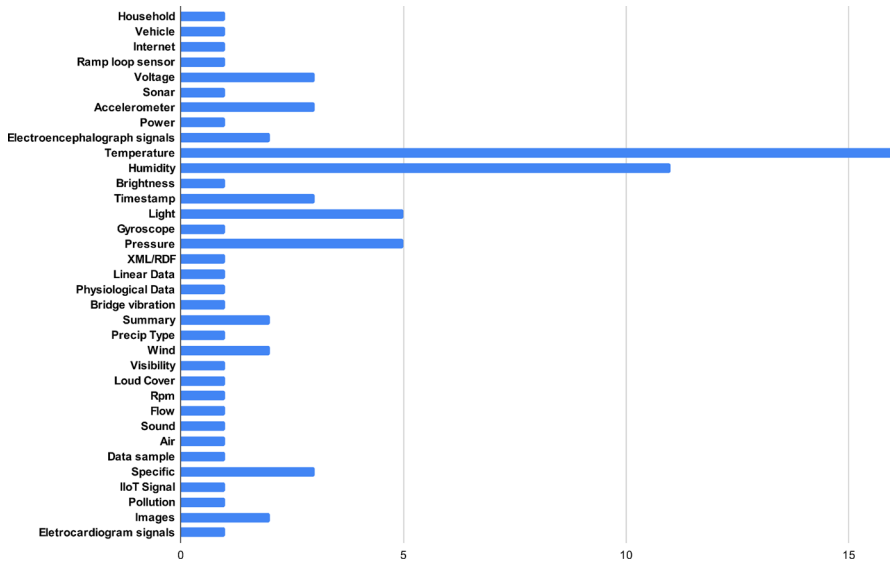
Fig. 7 Experimental data type

[37] a multi-layered data compression framework that performs an initial reduction phase in the fog layer and then ends the reduction in the cloud was proposed. Their solution begins reducing the fog layer through an initial compression of the collected data from the edge nodes. Then, they send the data to the cloud, where it performs a second compression process. The work [38] proposed a data collection approach performed at the edge layer that reduces the quantity of transmitted IoT data. They begin applying a data compression algorithm to the collected data before transmitting it to the cloud. After that, they rebuild these transmitted data on an edge node applying a ML technique.

4.3 Experiment data type

Figure 7 addresses the question *(MQ3: Which data type researchers are using for experimentation?)*. We verify that more than half studies used the temperature data stream in their experiments. Some studies that use this data type are: [22–24,39–42].

Temperature, humidity, pressure, and light were the most used data stream by the researchers in their experiments. We suggest that the usage is inherently related to some data characteristics. The first one is the context of the study that might some inferences to be introduced. IoT domain deals with a considerable variety of stream data due to the diversity of the sensors applications. Smart cities and smart homes, for instance, employ a wide variety of applications to monitor the weather and structural health. Because of that, the most used weather sensor (e.g., temperature, pressure, humidity, and light) are essential and widely used in such context. In [22] the authors advocate, for instance, that there is a correlation between pressure and temperature data. Therefore, the authors affirm that it is possible to infer the temperature data
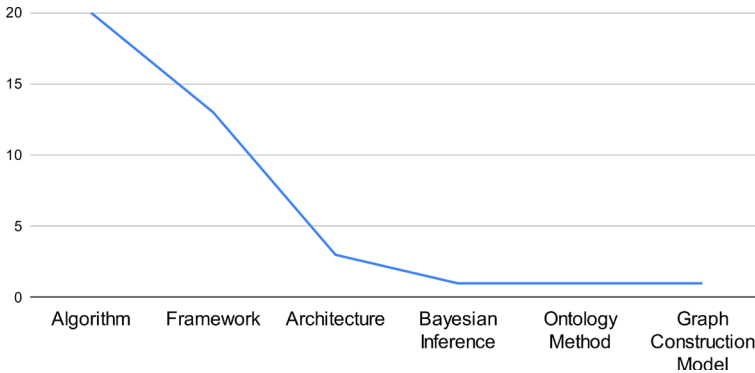
Fig. 8 Authors contributions

from pressure and vice-versa. The second characteristic is the usage of the datasets made available by [43] where 54 sensors are deployed in the Intel Berkeley Research lab. They have used Mica2Dot sensors together with weatherboards that collect time-stamped data information along with humidity, temperature, light, and voltage values. Many authors in this study used this dataset [22–24,32,33,40,42], and [44]. There are also cases where the used data is inherent to the domain. For instance, the work [45] uses a Kalman filter to minimize the data packet. They evaluated their results using standard intrusion detection datasets (e.g., NSL-KDD, KYOTO 2006+, CICIDS2017, and CICIDS2018) for instance. In such a case, they received a *"specific"* classification.

4.4 Author's contribution

Figure 8 addresses the question *(MQ4: Which was the contributed object proposed by the researchers?)*.

The main concepts of the authors' contributions are presented by trying to reach data reduction at the edge in a diverse range of problems and domains. It is important to note that some researches propose more than one algorithm contribution. Thus, the number of contributions presented can be different from the total of accepted studies. For instance, the authors [24] presented a two-tier data reduction approach with different algorithms solutions. The authors apply data compression algorithms at the first sensor nodes tier (e.g., delta encoding, RLE encoding). After, in the network gateways, a clustering algorithm is performed. In such a case, we count twice the algorithm solutions, and it might be different from the quantity number of accepted results. Moreover, this handling can be slightly confusing due to the interpretation bias that the classification term might present. For instance, which is the difference between a scheme, a method, and an approach? Even though we presented some definitions to classify the researchers' proposals in this study, our definition would introduce an interpretation bias. As we avoid this kind of issue in reducing interpretation bias, the original author term was used to condense and present the results.

Many of the presented researches propose a solution that is composed of more than one data reduction algorithm. For instance, the authors in [42] proposed a two-

level protocol called "Energy-efficient Data Transmission and Aggregation Protocol (EDaTAP)". Their solution is based on the fog computing paradigm, which is composed of two levels and performs its tasks both in the sensor device and at the fog gateway. An encoding algorithm is implemented at the sensor level to decrease the gathered sensed data volume and avoid energy wasting. In their solution, a clustering approach based on Dynamic Time Warping (DTW) is implemented at the fog smart gateway. The sensor devices send the measured data to the fog after the sensor algorithm processes it. Another processing algorithm is performed at the fog level to send the resulting data to the cloud for further analysis. In the work, [32] a multi-tier data reduction solution was implemented in the gateway and at the edge was also proposed. In addition, a data filtering algorithm based on the PIP technique was introduced in the gateway tear. On the other hand, at the edge, it was used a data fusion algorithm.

The studies [46] and [5] proposed a framework aiming to implement data reduction at the edge and the gateway of the network, respectively. In the first, the DISMISS framework was presented. It focuses on noise reduction in which imputing missing values and filtering stream data at the edge are reached. In the gateway, they applied the binning method to reduce the noise of raw data. To reach the missing values, they used a Dynamic Time Warping (DTW) similarity technique together with the PIP method to select the critical points. In the second, the authors proposed a data handler framework to reduce the data volume at the edge. Their solution comprises many data reduction methods (e.g., Data Sampling, Piecewise Approximation, Selective Forwarding, Perceptually Important Points, and Change Detection).

4.5 Data reduction implementation layer

Figure 9 address the *(MQ5:Where are the authors implementing the data reduction solutions?)* question. The authors' contributions can be implemented in different layers. It is worthy to notice that we did not consider a solution when it was proposed to perform data reduction only in the sensor device or in the cloud layer respecting the IC1 presented in Sect. 3.3.1. However, when the proposed solution is composed of many implementation layers, and one of these layers is the edge, we considered it in this work.

For instance, the authors [37] proposed a multi-layered data compression framework that performs an initial reduction phase in the fog layer and then ends the reduction in the cloud. The researchers in [22] presented a solution that uses Bayesian Inference Approach (BIA) both in the sensor device and in the gateway. In [24] the authors presented a two-tier data reduction approach with different algorithms solutions that use the sensor nodes tier and the network gateways to perform a clustering technique.

It is possible to visualize in Figure 9 that 17.1 % of the proposed data reduction solutions were deployed in one more layer in addition to the edge. It was referred to as a *"Two-level"* implemented layer. Data reduction deployed only at the edge (e.g., gateway, edge, fog, etc.) corresponds to the other 82.9% of all the proposed solutions, and it was referred to as a *"One-level"* implemented layer. These results were expected due to the research's objective, which is to present the data reduction solutions performed at the edge. An interesting finding here relates to the hybrid
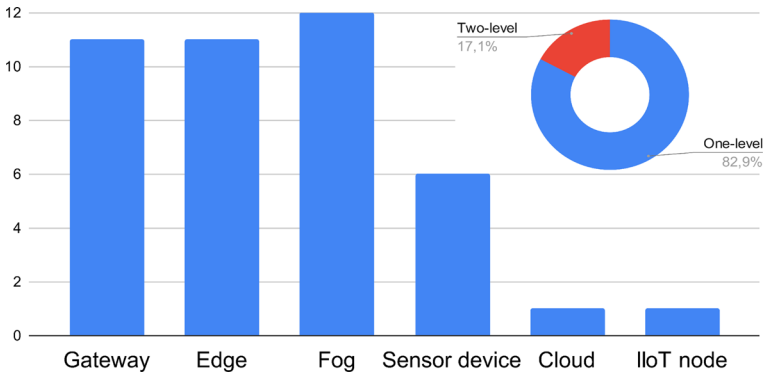
Fig. 9 Characteristics of the data reduction layers

solutions that use more than the edge layer to complement their proposals, even when we are interested in edge solutions.

Also in the Fig. 9, we observe the computational layer place of each data reduction solution. The majority of the methods were introduced at the fog level, followed by edge and gateway nodes. Although we understand that some of these layers might be synonymous, presenting their real place might bring many insights into this study. For instance, the solution [47] proposed a data reduction algorithm considering the industrial Internet of Things (IIoT) context. An IIoT node is introduced on a "Permanent Magnet Synchronous Motor (PMSM)" that reads two kinds of signals (leakage flux and vibration signals) through magnetic sensors and accelerometers without accessing the motor control system. A mechanical system accelerometer reads the motor's vibration signals and sends them to a constructed algorithm that filters these magnetic signals to reduce data. Although we highlighted the algorithm to extract the magnetic signal, the author also proposed other algorithms to handle signal mixing, separation, and rotation angle calculation.

4.6 Edge technology

Figure 10 addresses the question *(MQ6: Which hardware technologies were used by authors' to perform their data reduction solutions?)*.

First, we verified that the majority of the authors have used/constructed simulators to perform their experimentation. The works [42] and [24] have used a component-based C++ and modular *OMNeT++* [48] network simulator library to perform their experiments. The works [32,41,49] and [33] also used the same approach to evaluated their solution. They have written their simulator using the Python programming language and used the dataset presented by the work [43] to perform their experiments. The authors in [5] have simulated the environment in a Virtual Box Virtual Machine (VM). They simulated the capabilities of a lightweight gateway Raspberry Pi, a M2M Gateway, and a PC all using the VM. On the other hand, the works [40] and [50] simulated using the Java programming language. Finally, the works [23] and [45] used MATLAB to simulate the environment which simulates the edge technology.
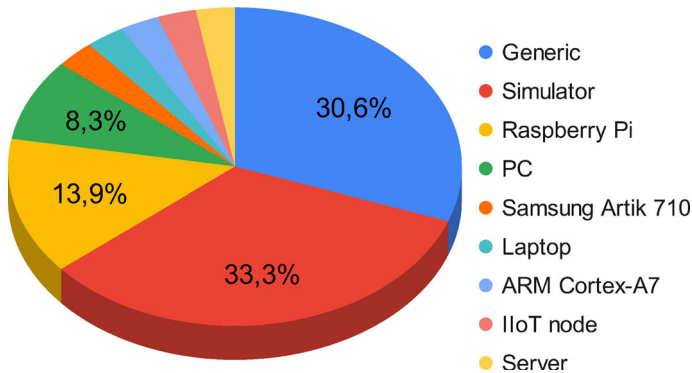
Fig. 10 Edge technology

Next, we verified that a generic edge technology was proposed when performing their experimentation. It means that the characteristics of the environment do not have much importance in the study. For instance, many of the proposed data reduction methods can be measured and experimented using only mathematical features. In [51] the authors evaluated the security and the performance of their data collection and computation offloading scheme by numerical results. To analyze their framework, in [31] the author's employed different test cases to verify its accuracy through similarities between the reduced and original datasets. In other existing works, this information was omitted. The authors [52] proposed an algorithm based on a pattern system that comprises a library and a classifier that handles bridge vibration sensors. The algorithm classifies the time-series data intervals into patterns that learn from set models. The data intervals are predicted and classified by a pattern classifier that returns the pattern of the interval data.

In some cases, the author [53] does not specify the used edge technology. In such a case, the used 10 edges nodes were classified as a *"generic"* edge technology due to the author's non-specification. They also used artificial datasets (e.g., Swiss-Roll, S-Curve, and S-Sphere). Other authors, [38] and [22], have used personal computers as an edge technology. Due to the limited quantity of pages, we summarized all selected studies that were used to answer the six research questions through a table available on this link.[2]

5 Threats to validity

As in any empirical experiment, this manuscript suffers threats to validity. Some of these elements are discussed and considered at this stage. In this section, we investigate and discuss the main risks this work presents.

The basis of this research aims to identify studies that have contributed to implementing data reduction solutions at the edge of IoT systems. It is worthy to notice that

---

[2] https://docs.google.com/spreadsheets/d/1A21KEEeS9_SEmERRJ6aKOr6mC-khA7v1e0ARbyTnyYg/edit?usp=sharing.

the accepted researches ought necessarily to have the data reduction concept as the first goal of the studies instead of another one. For instance, approaches that have the primary goal of improving data analytic at the edge and applying data reduction to reach the first goal were not accepted in this study. On the other hand, approaches that propose a data reduction method and used it to perform data analytic were considered in this mapping. Another interesting observation we should mention here is the used term by the researcher to refer to some process, which might include some data reduction technique. Some of them used the following terms *"data management", "data analytics", "energy saving", "real-time data processing", and "power consumption"*. To avoid introducing bias into the study, when these terms were identified, the articles were rejected.

6 Open research issues

Analyzing the studies and results we have presented in this mapping, a considerable quantity of insights and outcomes can be delivered considering the employment of the edge in IoT architectures. The use of an intermediate processing layer can be employed by many domains, providing an immense opportunity for new discovering and explorations. However, some specific issues and opportunities arise. Below, we discuss the employment of the processing layer in IoT technologies, considering some topics for research investigations and possible limitations of its usability.

6.1 Latency and scalability

Introducing a preprocessing layer in an IoT infrastructure might bring many advantages concerning data quality. Using an edge layer to filter and reduce transmitted data might improve the network bandwidth, energy, I/O throughput, cloud storage, and traffic costs. However, depending on the selected architectural configuration, the latency can be reduced. In the work [5], the authors discussed the introduced latency of the forwarded data to the cloud. In their solution many data reduction methods were used, although only the PIP data reduction method was considered for latency analyses. Depending on the layer the data reduction solution is introduced, the latency might be bigger or smaller. As far as the data storage is from its generated point, the bigger the latency will be introduced. The authors solutions, [22,23] and [24] implemented data reduction directly in the sensor layer. It is expected that it might reduce the quantity of data sent to the edge to be preprocessed, reducing then the latency in the applications layer. Implementing ML near the data production layer also tends to improve latency, as unnecessary data tends to be ignored, leaving the network free for meaningful data.

6.2 Artificial intelligence

The usage of AI to support data preprocessing seems to be an interesting solution to increase data quality when using the edge architecture model. However, depending on the selected architectural schema to implement the solution, the AI module might

suffer from inaccuracy in data results. Moving the ML closer to the data source might reduce the data transmission. The authors in [18] proposed an architecture that introduces the ML closer to the data source. They discuss the three main techniques (e.g., Device and Edge, Edge and Cloud, and Device and Cloud) to perform ML considering the edge computing paradigm. Researchers in [21] proposed a framework that uses autoencoders model at the edge to apply data compression while in [19] the encoder part of the autoencoder was deployed in the edge while the decoder in the cloud.

6.3 Edge hardware

Employing the data reduction solutions on edge has proved its importance over the years. However, depending on the faced problem and domains, some edge solutions seemed to be more suitable than others. For instance, solutions for the Industrial Internet of Things (IIoT) domain need attention regarding hardware selection. In [47] the authors introduced the "IIoT node on a Permanent Magnet Synchronous Motor (PMSM)". This node was constructed to collect the vibration signals for specific motors conditions verification reading different magnetic sensors signals. In the [5] work, the authors pointed that the cache's size does not imply better effectiveness of the forwarded data. On the other hand, there are also cases in which all the network components, including the edge hardware, were simulated. The usage of simulators to experiment and find the better hardware to use as an edge device seems to be an interesting solution. In [42] and [24] works, the authors used the *OMNeT++* [48] network simulator library, while the works [23] and [45] used MATLAB. The authors [32,41,49] and [33] wrote their simulations using the Python programming language, while the works [40] and [50] simulated using the Java programming language. Finally, the selection of the edge hardware when implementing the data reduction techniques in the edge-based architecture can be explored considering different data workloads classification.

7 Conclusions

Internet of Things technology connects plenty of devices from different types and characteristics through a network, allowing communications between these devices and systems. The large quantity of the generated data sources, processed, and stored has increased considerably, imposing some challenges in the process of making that data available. To solve such an issue, data reduction and preprocessing are promising concepts that help to manage this data efficiently. Applying data reduction, considering the edge computing paradigm, that handles data closer to the generated point before being sent to the cloud, has emerged as an efficient solution. The gateways designs were projected according to the features and requirements of the different applications on services goal to perform this task. This study investigates the data reduction solutions performed exclusively at the edge of IoT systems. To reach this objective, we performed a Systematic Literature Mapping (SLM), which allows us to answer six promissory questions in such context. The main findings of this study relate to the

answer to the investigated questions that justify this mapping. The progressing time of this research field, the used techniques to reduce data at the edge, the used data type in experimentation, the authors proposed solutions of data reduction, the architectural data reduction layer implementation, and the hardware technologies used to perform data reduction at the edge were some investigated questions.

References

1. Ashton K et al (2009) That 'internet of things' thing. RFID J 22(7):97–114
2. Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. Comput Netw 54(15):2787–2805
3. Dilley J, Maggs B, Parikh J, Prokop H, Sitaraman Ramesh, Weihl Bill (2002) Globally distributed content delivery. IEEE Internet Comput 6(5):50–58
4. IBM (2021) Data reduction stages
5. Papageorgiou A, Cheng B, Kovacs E (2015) Real-time data reduction at the network edge of internet-of-things systems. In: *2015 11th international conference on network and service management (CNSM)*, pp. 284–291. IEEE
6. ur Rehman MH, Liew CS, Abbas A, Jayaraman PP, Wah TY, Khan SU, (2016) Big data reduction methods: a survey. Data Sci Eng 1(4):265–284
7. Guan Z, Tongkai JX, Qian YM, Hong X (2017) A survey on big data pre-processing. In *2017 5th Intl Conf on Applied Computing and Information Technology/4th Intl Conf on Computational Science/Intelligence and Applied Informatics/2nd Intl Conf on Big Data, Cloud Comput Data Sci (ACIT-CSII-BCD)*, pp. 241–247. IEEE
8. Hansong Xu, Wei Yu, Griffith D, Golmie Nada (2018) A survey on industrial internet of things: a cyber-physical systems perspective. IEEE Access 6:78238–78259
9. Cook AA, Mısırlı Göksel, Fan Z (2019) Anomaly detection for iot time-series data: A survey. IEEE Internet Things J 7(7):6481–6494
10. Ding W, Jing X, Yan Z, Yang LT (2019) A survey on data fusion in internet of things: towards secure and privacy-preserving fusion. Inform Fusion 51:129–144
11. Khan Latif U, Yaqoob Ibrar, Tran NH, Ahsan Kazmi SM (2020) Tri Nguyen Dang, and Choong Seon Hong. Edge-computing-enabled smart cities: a comprehensive survey. IEEE Internet Things J 7(10):10200–10232
12. Dehkordi SA, Farajzadeh K, Rezazadeh J, Farahbakhsh R, Sandrasegaran K, Dehkordi MA (2020) A survey on data aggregation techniques in iot sensor networks. Wireless Networks, 26(2):1243–1263
13. Abdulzahra SA, M Al-Qurabat AK (2021) Data aggregation mechanisms in wireless sensor networks of iot: a survey. Int. J. Comput. Dig. Syst
14. Jane VA et al (2021) Survey on iot data preprocessing. Turkish J Comput Math Educ (TURCOMAT) 12(9):238–244
15. Li PH, Youn HY (2020) Gradient-based adaptive modeling for iot data transmission reduction. Wireless Netw 26(8):6175–6188
16. Salim C, Makhoul A, Couturier Raphaël (2020) Energy-efficient secured data reduction technique using image difference function in wireless video sensor networks. Multimed Tools and Appl 79(3):1801–1819
17. M Al-Qurabat AK, Mohammed ZA, Hussein ZJ (2021) Data traffic management based on compression and mdl techniques for smart agriculture in iot. Wireless Personal Communications, pp. 1–32
18. Hafeez T, Lina X, Mcardle Gavin (2021) Edge intelligence for data handling and predictive maintenance in iiot. IEEE Access 9:49355–49371
19. Ghosh AM, Grolinger K (2019) Deep learning: Edge-cloud data analytics for iot. In: 2019 IEEE Can Conf Electric Comput Eng (CCECE), pp. 1–7. IEEE
20. Suryadevara NK (2021) Energy and latency reductions at the fog gateway using a machine learning classifier. Sustaine Comput: Inform Syst, p. 100582
21. Ndubuaku MU, Ali MK, Anjum A, Liotta A, Reiff-Marganiec S (2020) Edge-enhanced analytics via latent space dimensionality reduction. In: 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), pp. 87–95. IEEE

22. Razafimandimby C, Loscri V, Vegni AM, Aourir D, Neri A (2017) A bayesian approach for an efficient data reduction in iot. In :Interoperability, Safety and Security in IoT, pp. 3–10. Springer
23. Fathalla A, Li K, Salah A, Mohamed MF (2021) An lstm-based distributed scheme for data transmission reduction of iot systems. Neurocomputing
24. M Al-Qurabat AK, Jaoude CA, Idrees AK (2019) Two tier data reduction technique for reducing data transmission in iot sensors. In: 2019 15th Int Wireless Commun & Mobile Comput Conf (IWCMC), pp. 168–173. IEEE
25. Kitchenham BA, Budgen D, Brereton OP (2011) Using mapping studies as the basis for further research-a participant-observer case study. Inform Softw Technol 53(6):638–651
26. Keele S et al. (2007) Guidelines for performing systematic literature reviews in software engineering. Technical report, Technical report, Ver. 2.3 EBSE Technical Report. EBSE
27. Petticrew M, Roberts H (2006) Systematic reviews in the social sciences: a practical guide. Blackwell Publishing CrossRef Google Scholar, Malden USA
28. Budgen D, Brereton P (2006) Performing systematic literature reviews in software engineering. In: Proceedings of the 28th international conference on Software engineering, pp. 1051–1052. ACM
29. Ndiaye M, Oyewobi SS, Abu-Mahfouz AM, Hancke GP, Kurien AM, Djouani K (2020) Iot in the wake of covid-19: A survey on contributions, challenges and evolution. IEEE. Access 8:186821–186839
30. Yunkang X, Kishi T (2018) An ontology-based iot communication data reduction method. In: 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 321–325. IEEE
31. Ling WS, Yaik OB, Yue LS (2017) A novel data reduction technique with fault-tolerance for internet-of-things. In: Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing. pp 1–7
32. Feng L, Kortoçi P, Liu Y (2017) A multi-tier data reduction mechanism for iot sensors. In: Proceedings of the Seventh International Conference on the Internet of Things. pp 1–8
33. Ismael WM, Gao M, Al-Shargabi AA, Zahary A (2019) An in-networking double-layered data reduction for internet of things (iot). Sensors 19(4):795
34. Sari K, Riasetiawan M (2018) The implementation of timestamp, bitmap and rake algorithm on data compression and data transmission from iot to cloud. In: 2018 4th International Conference on Science and Technology (ICST), pp. 1–6. IEEE
35. Burrello A, Marchioni A, Brunelli D, Benini L (2019) Embedding principal component analysis for data reduction in structural health monitoring on low-cost iot gateways. In: Proceedings of the 16th ACM International Conference on Computing Frontiers. pp. 235–239
36. Chang K-C, Chiang M-H (2019) Design of data reduction approach for aiot on embedded edge node. In: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), pp. 899–900. IEEE
37. Hossain K, Roy S (2018) A data compression and storage optimization framework for iot sensor data in cloud storage. In: 2018 21st International Conference of Computer and Information Technology (ICCIT), pp. 1–6. IEEE
38. Azar Joseph, Makhoul Abdallah, Barhamgi Mahmoud, Couturier Raphaël (2019) An energy efficient iot data compression approach for edge machine learning. Futur Gener Comput Syst 96:168–175
39. Alahmed A, Alrasheedi A, Alharbi M, Alrebdi N, Aleasa M, Moulahi T (2020) Machine learning for real-time data reduction in cloud of things. In: 2020 2nd International Conference on Computer and Information Sciences (ICCIS), pp. 1–6. IEEE
40. Harb H, Jaoude CA, Makhoul A (2020) An energy-efficient data prediction and processing approach for the internet of things and sensing based applications. Peer-to-Peer Netw Appl 13(3):780–795
41. Ismael WM, Gao M, Yemeni Zaid (2021) Esrra-iot: Edge-based spatial redundancy reduction approach for internet of things. Internet of Things 14:100388
42. Idrees AK, Al-Qurabat M (2021) Energy-efficient data transmission and aggregation protocol in periodic sensor networks based fog computing. J Netw Syst Manag 29(1):1–24
43. Bodik Peter, Hong Wei (2004) Carlos Guestrin. Mark Paskin, and Romain Thibaux. Intel lab data. Online dataset, Sam Madden
44. Tianqi Y, Wang X, Shami A (2017) A novel fog computing enabled temporal data reduction scheme in iot systems. In: GLOBECOM 2017-2017 IEEE Global Communications Conference, pp. 1–5. IEEE
45. Gavel S, Raghuvanshi AS, Tiwari S (2021) Distributed intrusion detection scheme using dual-axis dimensionality reduction for internet of things (iot). The J Supercomput pp. 1–24
46. Balasubramanian S, Meyyappan T (2019) Enhancing the computational intelligence of smart fog gateway with boundary-constrained dynamic time warping based imputation and data reduction. In:

2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC), pp. 15–23. IEEE

47. Wang X, Siliang L, Huang W, Wang Q, Zhang S, Xia M (2021) Efficient data reduction at the edge of industrial internet of things for pmsm bearing fault diagnosis. IEEE Trans Instrum Meas 70:1–12
48. Varga A (2010) Omnet++. In Modeling and tools for network simulation. Springer, pp. 35–59
49. Idrees SK, Idrees AK (2021) New fog computing enabled lossless eeg data compression scheme in iot networks. J Ambient Intell Humanized Comput pp. 1–14
50. Papageorgiou A, Cheng B, Kovacs E (2015) Reconstructability-aware filtering and forwarding of time series data in internet-of-things architectures. pp 576–583
51. Chen S, Zhu X, Zhang H, Zhao C, Yang G, Wang K (2020) Efficient privacy preserving data collection and computation offloading for fog-assisted iot. IEEE Trans Sustain Comput 5(4):526–540
52. Chen A, Liu F-H, Wang S-D (2019) Data reduction for real-time bridge vibration data on edge. In: 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 602–603. IEEE
53. Li G, Cao J, Junhua W, Ren X, Haili Y (2020) Dimension reduction algorithm based on adaptive maximum linear neighborhood selection in edge computing. IEEE Internet of Things Journal