A Comprehensive Survey on Collaborative Data-access Enablers in the IIoT

DANFENG SUN, JUNJIE HU, and HUIFENG WU, Key Laboratory of Discrete Industrial Internet of Things of Zhejiang Province, Hangzhou Dianzi University, China
JIA WU, JIAN YANG, and QUAN Z. SHENG, School of Computing, Macquarie University, Australia
SCHAHRAM DUSTDAR, Distributed Systems Group, TU Wien, Austria

The scope of the Industrial Internet of Things (IIoT) has stretched beyond manufacturing to include energy, healthcare, transportation, and all that tomorrow's smart cities will entail. The realm of IIoT includes smart sensors, actuators, programmable logic controllers, distributed control systems (DCS), embedded devices, supervisory control, and data acquisition systems—all produced by manufacturers for different purposes and with different data structures and formats; designed according to different standards and made to follow different protocols. In this sea of incompatibility, how can we flexibly acquire these heterogeneous data, and how can we uniformly structure them to suit thousands of different applications? In this article, we survey the four pillars of information science that enable collaborative data access in an IIoT—standardization, data acquisition, data fusion, and scalable architecture—to provide an up-to-date audit of current research in the field. Here, standardization in IIoT relies on standards and technologies to make things communicative; data acquisition attempts to transparently collect data through plug-and-play architectures, reconfigurable schemes, or hardware expansion; data fusion refers to the techniques and strategies for overcoming heterogeneity in data formats and sources; and scalable architecture provides basic techniques to support heterogeneous requirements. The article also concludes with an overview of the frontier researches and emerging technologies for supporting or challenging data access from the aspects of 5G, machine learning, blockchain, and semantic web.

CCS Concepts: • **Information systems → Data management systems**; • **Computer systems organization → Cloud computing**; **Data flow architectures;**

ACM Reference format:
Danfeng Sun, Junjie Hu, Huifeng Wu, Jia Wu, Jian Yang, Quan Z. Sheng, and Schahram Dustdar. 2023. A Comprehensive Survey on Collaborative Data-access Enablers in the IIoT. *ACM Comput. Surv.* 56, 2, Article 50 (September 2023), 37 pages.
https://doi.org/10.1145/3612918

This work was supported in part by a grant from the National Key R&D Program of China (2022YFB3304600), in part by a grant from the National Natural Science Foundation of China (No. U21A20484), and in part by a grant from the Science and Technology Program of Zhejiang Province (No. 2022C01016).
Authors' addresses: D. Sun, J. Hu, and H. Wu (Corresponding author), Key Laboratory of Discrete Industrial Internet of Things of Zhejiang Province, Hangzhou Dianzi University, No. 1158 Erhao Road, Hangzhou, China; emails: {danfeng.sun, hujj, whf}@hdu.edu.cn; J. Wu, J. Yang, and Q. Z. Sheng, School of Computing, Macquarie University, Balaclava Road, Sydney, Australia, 201101; emails: {jia.wu, jian.yang, michael.sheng}@mq.edu.au; S. Dustdar, Distributed Systems Group, TU Wien, Balaclava Road, Vienna, Austria, 1040; email: dustdar@dsg.tuwien.ac.at.

0360-0300/2023/09-ART50 $15.00
https://doi.org/10.1145/3612918

1 INTRODUCTION

The **Industrial Internet of Things (IIoT)** is a significant and powerful driving force behind the fourth industrial revolution. Through a massive number of heterogeneous sensing devices, each with the ability to calculate, communicate, and collaborate, IIoT is bringing interconnectivity and intelligence to industrial systems. For engineers and data scientists, IIoT is an exciting, challenging, and inescapable aspect of research. Figure 1 illustrates the breadth of sectors IIoT is benefiting, ranging from smart factories and logistics to smart cities and transportation. Table 1 summarizes the many surveys that have been undertaken on different aspects of IIoT.

Viewed as the fourth revolution in manufacturing, intelligent manufacturing is at the forefront of IIoT. Through integration and interoperability, factories are becoming smart. Devices such as **programmable logic controllers (PLC)** and embedded controllers are being combined with sensor networks and cloud-based control systems [267]. This enables data exchanging and knowledge sharing between two systems [146], as well as seamless operations across organizational boundaries.

IIoT is also making energy production, delivery, and usage more efficient. With the benefit of smart sensors, network devices, and data analytics, energy companies are able to monitor asset performance and investigate accidents remotely and safely [11, 58, 61, 148].

In healthcare, IIoT is showing enormous promise as a way for medical professionals to monitor, aggregate, and act on patients' medical data in real-time [86, 123]. IIoT is also enabling demand-based production lines for factory and laboratory processes like drug packing [228]. In logistics and supply chains, IIoT-enabled technologies mean products and materials can be tracked in real-time. This tracking data help enterprises to manage their warehouses and transport fleets, as well as optimize their processes [24, 237].

In terms of transportation, IIoT is improving vehicle-to-vehicle communications and remote monitoring for smart cars [125], smart trains [65], **unmanned aerial vehicles (UAVs)** [10], and so on. In addition to networking, these applications are helping to reduce maintenance costs, fuel consumption, and accident response times. However, the shining jewel of IIoT is the smart city, where physical infrastructure, services, and networking all integrate to make almost every element of a city more intelligent and more efficient [255]. As just a few examples, IIoT is making possible: structural health monitoring, pipeline network monitoring, camera surveillance networks, urban traffic monitoring, smart grid monitoring, and streetlight monitoring [54].

Importantly, it is not just academic research into IIoT that is thriving. Many commercial enterprises are also building IIoT platforms to capitalize on this new revolution, both horizontal and vertical. Vertical platforms focus on providing a suite of services to support an entire field from the ground up. The success of these types of platforms often depends heavily on the host's competitive advantages and experience in the field. Examples include MindSphere [206] from Siemens, EcoStructure from Schneider Electric [197], and ABB Ability [5].

Horizontal platforms are a more ambitious endeavor, providing more generalized applications to suit the requirements across a range of industries. At this stage, these platforms are mainly driven by huge cloud service providers, such as the AWS IoT [13] by Amazon and Azure IoT [161] by Microsoft. The drawback of horizontal platforms is that many **small and medium-sized enterprises (SMEs)** are required to do outsourcing work due to their lack of specific industry experience. GE's Predix is a typical case of where a horizontal transformed into a vertical platform.

Indeed, the road ahead for IIoT is bright. However, as with all maturing technologies, there are still challenges to overcome. At present, IIoT is on the precipice of explosive growth. The four main issues slowing take-off are:

(1) The complexity of industrial plants and equipment. The sheer range of equipment and countless combinations between them turns digitizing any factory into a huge project.

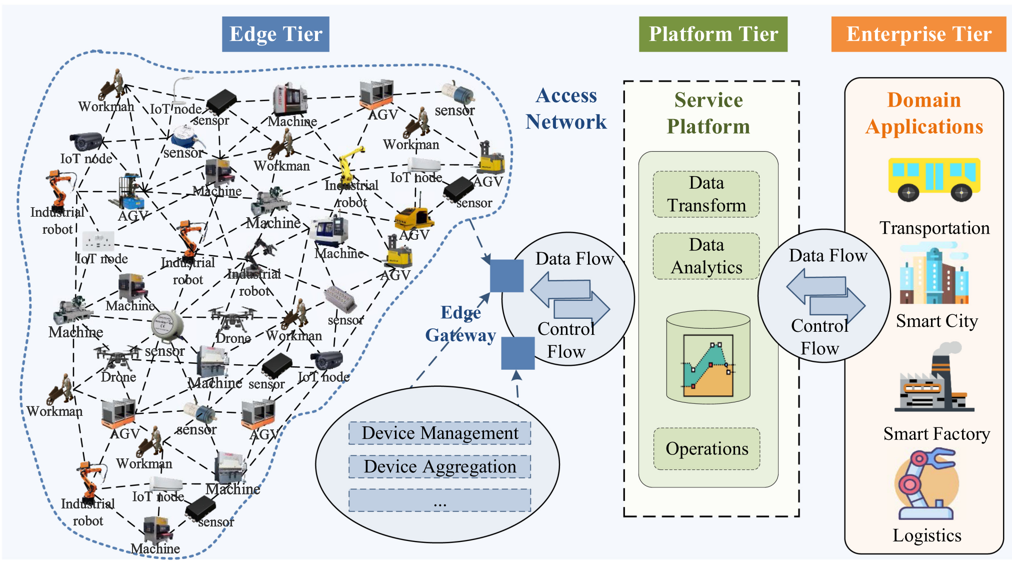

Fig. 1. Three-tier IIoT system architecture: (1) edge tier: monitoring, control; (2) platform tier: data collection and conversion; (3) enterprise tier: business and application domains (analytics, management).

Table 1. Surveys of IIoT Application Fields

Application Field	Reference	Contributions							Description
		O	D	F	C	T	K	A	
Intelligent manufacturing	Zheng et al. [267]	✓	–	–	✓	✓	–	✓	Intelligent manufacturing, or smart factory, is viewed as the fourth revolution in the manufacturing industry.
	Lu et al. [146]	–	✓	–	✓	✓	–	✓	
	Mabkhot et al. [152]	✓	✓	✓	✓	–	✓	✓	
	Cheng et al. [40]	✓	–	✓	–	✓	–	✓	
Energy	Fang et al. [61]	✓	–	–	–	–	✓	✓	Energy production, delivery, and use are getting more efficient with IIoT.
	Alahakoon et al. [11]	✓	✓	–	✓	✓	✓	✓	
	Lund et al. [148]	–	✓	–	✓	✓	✓	✓	
	Erol-Kantarci et al. [58]	–	✓	–	✓	–	✓	✓	
Healthcare	Hossain et al. [86]	✓	–	–	✓	–	✓	✓	Medical processes are becoming safer through better monitoring, and demand-based production of drugs is becoming possible.
	Wan et al. [228]	✓	–	–	✓	✓	–	✓	
	Kim et al. [123]	–	–	–	✓	✓	–	✓	
	Yue et al. [257]	✓	✓	✓	✓	✓	–	–	
Logistics and supply chain	Barreto et al. [24]	✓	✓	✓	✓	✓	✓	✓	Materials and products can be tracked and reported.
	Witkowski et al. [237]	✓	✓	–	–	✓	✓	✓	
Smart cities	Yin et al. [255]	✓	✓	–	–	✓	✓	–	Joining physical infrastructure and services with information and communications technology to make cities more intelligent and efficient.
	Petrolo et al. [177]	✓	✓	–	–	✓	✓	✓	
	Gharaibeh et al. [69]	✓	✓	✓	✓	–	✓	✓	
	Du et al. [54]	–	✓	–	✓	✓	✓	✓	
	Lau et al. [133]	–	–	–	✓	✓	–	✓	
Transportation	Fraga-Lamas et al. [65]	–	–	–	–	–	–	–	Vehicle-to-vehicle communications and remote monitoring.
	Rathore et al. [185]	✓	✓	–	✓	✓	✓	✓	
	Ding et al. [51]	–	–	–	✓	✓	✓	✓	
	Giang et al. [70]	–	–	–	–	✓	✓	✓	
	Yan et al. [250]	–	✓	–	✓	✓	–	✓	
	Jan et al. [104]	✓	–	✓	✓	–	✓	✓	

O = Origin, D = Definition, F = Fundamentals, C = Challenges, T = Trends, K = Key Technologies, A = Applications.

(2) The heterogeneity of data, data sources, and data collection protocols. Every machine, equipment controller, and sensor manages its data in a different way, and although there are many conversion tools, exchange frameworks, and international standards for compatibility, the conventions are far from universal. Getting all the machines and their manufacturers to a point where implementation is transparently "plug and play" has long been a challenge facing the industry.

(3) The collecting data needs to be pre-processed to a required structure before it can be used for analytics. However, the edge devices cannot know the actual data source mapping to many dynamic applications, so it is impossible to pre-determine which data preprocessing program the edge device should apply. A method of quickly constructing a data pre-processing logic according to the needs of the scenario at hand is urgently needed, as is a strategy for deploying diverse and personalized intelligent services for different edge devices.
(4) Traditional computing architectures are straining. A few years ago, the expectation was that all IoT would move toward cloud-based data centers, and those cloud centers would be sufficiently powerful to handle all IoT needs [187]. However, while many consumer-level IoT systems can benefit from this scheme, most end-point devices in an IIoT application demand much lower latency than a central server can provide. In industrial applications, computing generally needs to happen much closer to the sensor or device producing the data. Edge computing is an emerging computing architecture that allows for this [33, 203, 238], but more solutions that integrate cloud computing and edge computing need to be designed and applied [68]. Compared to cloud-only solutions, hybrid approaches can provide more efficient and flexible computing with lower latency and high resource utilization to support the fast decision and response times required for industrial applications. In addition, scalability is critical to IIoT architectures, as the exponential growth of data generated by connected devices and industry requires scalable resource management and deployment technologies that can scale horizontally to meet customer needs.

Around these issues, we investigated many past surveys and research articles. We list the high-quality surveys from 2018 in Table 2 and separated them into several research directions as their authors claimed, namely, edge computing, identity resolution system, interoperability, network slicing, security, privacy, blockchain, digital twin, deep learning, 5G, fault diagnosis, and additive manufacturing. Except for a few earlier review articles [122, 208] using a comprehensive perspective, other surveys provide quality introductions to their specific directions and throughout the state-of-the-art literature as well as the existing commercial platforms. We have an insight that the key to the above-mentioned issues is data. A more precise explanation is to focus on the enablers for the common stages of the data flow while optimizing the usage of edge-cloud resources and integrating emerging techniques. We conclude it as **collaborative data access (CDA)** that focuses on existing promising research of data awareness, data acquisition, data fusion, data security, and scalable architecture to alleviate the above issues, respectively. The state-of-the-art research lacks a comprehensive survey on the collaboration required across the five areas to address the issue of accessing massive and heterogeneous industrial data. Table 2 compares the difference between the enabled CDA and other recent surveys.

Hence, our focus is the current research into tools, methods, and technologies that enable CDA for massive industrial applications, especially large-scale edge networks. Through our investigation, we also aim to provide research directions for future studies in this field. Based on our investigation into CDA-related fields, we described the current research status using progressive statements of problems and corresponding solutions. We also identified unresolved issues, providing research directions for future studies in this field. We have divided the review into six sections, each representing a key scientific issue. These are standardization in the IIoT, data acquisition, data fusion, scalable architecture, and emerging fields, as shown in Figure 2. Before a thing in an IIoT system can send or receive data, it must be aware of what and where other things in the system are. The servers and edge devices must also be aware of how to process all the data produced by the end units. The process of awareness is standardization, which is the first step in enabling CDA. These issues are detailed in Section 2. Section 3 concerns data acquisition, which includes the methods used to connect to and communicate with entities in the network. Data fusion refers to the

Table 2. Comments on High-quality Surveys from the Perspective of IIoT Research Directions

Directions	Surveys and their relevance to the collaborative data access	Points of concern
Comprehensiveness	Sisinni et al. [208] compared the IoT, IIoT, Industry 4.0, and CPS in detail around interoperability, security and privacy, highlighting the opportunities and challenges. While exhibiting some overlap with the CDA, there is no explicit concentration on a specific topic.	• Energy efficiency • Real-time performance • Coexistence • Interoperability • Security and privacy
	Khan et al. [122] added the discussion of enabling technologies for IIoT compared with Reference [208], no explicit focus on a specific topic, but pointed out the challenge of the collaborative IIoT and involved blockchain and data fusion, which have some hints for the CDA.	• Cloud computing • Artificial Intelligence • Blockchain • Big Data • Augmented and virtual reality
Edge computing	Qiu et al. [183] discussed concepts, architectures, applications, challenges, and opportunities of edge computing under the context of the IIoT, which is different from the CDA, which emphasizes the process of data flow and leverages the cloud and edge resources.	• Routing • Task scheduling • Data storage and analytics • security • Standardization
Identity resolution system	Ren et al. [189] provided a comprehensive discussion on the identity resolution system, which is a part of the CDA first stage but is not centered on Collaboration.	• Systems • Comparison • Prospect
Interoperability	Hazra et al. [81] focused on interoperability, with no emphasis on the process of data flow, however, it provides hints for the CDA.	• Standards • Protocols • Models
Network slicing	Wu et al. [242] provided an investigation on network slicing from an application perspective, which is service-centric rather than data-centric.	• Smart transportation • Smart energy • Smart factory
Security	Figueroa et al. [63] detailed security strategies and issues of commonly used protocols and discussed the usage of the Common Vulnerability Scoring System, which is an important reference but is not the focus of CDA.	• Information technologies • Internet of Things • Operational technologies
	Tange et al. [222] comprehensively introduced the security requirements as CIA (i.e., Confidentiality, Integrity, Availability), security monitoring, maintainability, authentication, network security, access control, resilience, models and methodologies, and data security and sharing; and solutions with fog computing, which provides hints for securing the CDA but is not the focus of this survey.	• Quantitative analysis • Comparison between IoT and IIoT • Research method • Security requirements • Fog computing
Privacy	Jiang et al. [109] invested the differential privacy in the IIoT, a further topic after the CDA.	• Privacy protection • Differential privacy • Deep learning
Blockchain	Huo et al. [93] comprehensively discussed the blockchain in IIoT, where data collection and sharing provide hints for the CDA but no explicit focus on the whole process.	• Devices security • Data collection and sharing
Digital twin	Minerva et al. [162] comprehensively discussed the features, value, applications, and issues of the digital twin under the background of IIoT, where CDA can be its important base.	• Augmented and virtual reality • Multiagent systems • Virtualization • Features and value
Deep learning	Khalil et al. [121] presented the deep learning fields of predictive maintenance, real-time manufacturing, collaborative robotics, assets tracking, smart meters, healthcare monitoring, human resource, mining, agricultural, telecom, energy, transportation, waste management, advertisement, and petrochemical, which are not our focus.	• Concepts • Use cases • Security • Challenges
	De et al. [47] introduced the deep generative models for industrial applications in IIoT. Since data must be required for these models, CDA can be the base. Correspondingly, deep generative models can also empower the CDA.	• Deep generative models • Anomaly detection • Trust-boundary protection • Network traffic prediction • Platform monitoring
5G	Jiang et al. [111] introduced the 5G for IIoT, which is also important for CDA but is not the main focus of this survey.	• Environmental characteristics • Potential frequency bands • Channel model and parameters
Fault Diagnosis	Chi et al. [43] compared knowledge-based, plain model-based, and data-driven diagnosis approaches and discussed the future of decentralized knowledge-based fault diagnosis, where CDA can be one of its cornerstones.	• Ontology • Deductive/inductive reasoning • Decentralized implementations
Additive manufacturing	Haghnegahdar et al. [79] discussed the additive manufacturing enabled by IIoT, where CDA is its important base.	• Cloud computing • Cloud manufacturing • IIoT

techniques and strategies for overcoming heterogeneity in data formats and sources [119, 262]. Recent developments in this quarter are reviewed in Section 4. Scalable architecture, discussed in Section 5, is the scaffolding of CDA and customized applications in a scalable way. In addition to the pressing issues facing researchers, we have included a survey of emerging technologies in the field in Section 6. As potential accelerators of CDA, the topics include 5G, machine learning, semantic web, and blockchain. The article concludes in Section 7 with a brief summary of the topics covered.

2 STANDARDIZATION IN THE IIOT

Today, communication protocols, interfaces, and standards developed by many different equipment manufacturers are the cause of high complexity, and solving these problems is the main goal of different international initiatives. As just a few examples of these consortiums, in the United States, the **Industrial Internet Consortium (IIC)** is committed to achieving universal interconnection methods and intelligent analysis [44]. In China, the "Made in China 2025" plans to make a manufacturing revolution upon **information and communication technology (ICT)** [243]. In Germany, the **"Industry 4.0" (I4.0)** initiatives focus on the future of manufacturing through industrial ICT [198]. The main standardization result of I4.0 is its reference architecture model, known as RAMI 4.0 [234].

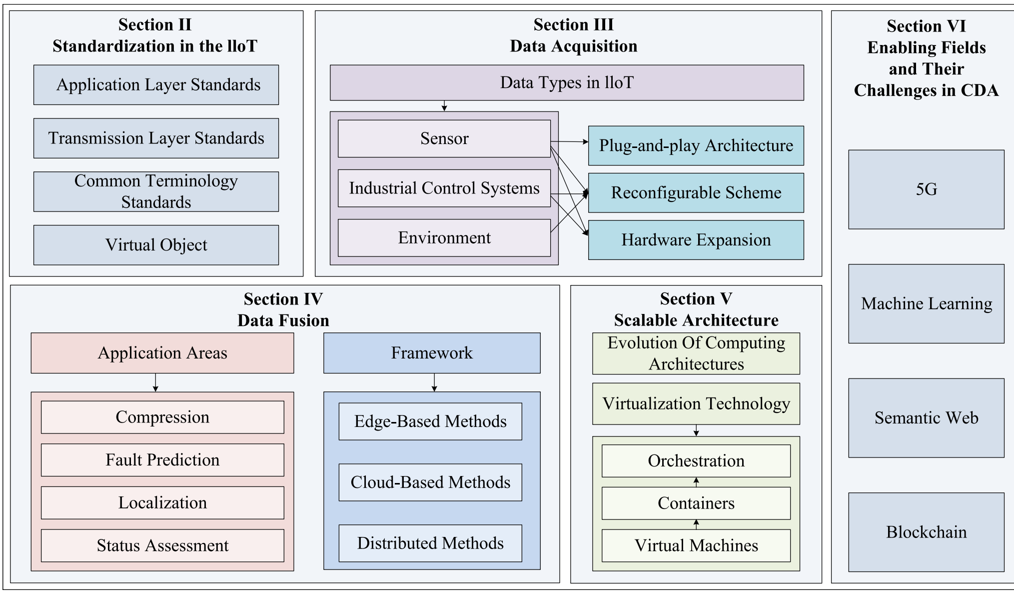

Fig. 2. The structure of this article.

To support the goals of different international initiatives, standardization is essential [176]. The technical specifications defined by the consensus are recognized standards that ensure uniformity and encourage interoperability, simplifying the work of stakeholders. Due to the wide variety of standards in IIoT, this article focuses on standards related to CDA. **International Electrotechnical Commission (IEC)** is the earliest organization participating in standardization work. IEC/TC65 is one of their groups, which works on measurement and process automation and is the core of the I4.0 International Standardization Technical Committee [147]. While Technical Committee 184 from **International Organization for Standardization (ISO)** makes efforts on automation systems and integration [1].

Virtualizing physical assets, as shown in Figure 3, is the basis for data access. This section introduces virtual objects and their common terminology standards. In addition, this article divides the application fields of IIoT general standards and technical standards into common terminology, perception layer, transmission layer, and application layer, which are mapped to the traditional IIoT structure [28].

2.1 Virtual Object

Describing highly heterogeneous physical assets is essential for better access to them, i.e., virtual representations of industrial assets. Standard data formats for describing virtual objects include eCl@ss or IEC **Common Data Dictionary (CDD)**, which can be specified by the standard IEC 61360 [259]. A typical example of a virtual interface application for physical assets is the **Asset Management Shell (AAS)** in I4.0. The following is an example of how to virtualize objects in I4.0:

Figure 3 shows the form of the AAS, which can be divided into **digital factories (DF)** Header and Body, and DF corresponds to the concept of DF framework in IEC 62832 [223]. DF Header includes properties that identify the physical asset and AAS itself as one uniquely I4.0 component [166]. This properties list is called a manifest, which is a set of defined meta-information that can be accessed from the outside [220]. DF Body inherits many basic models

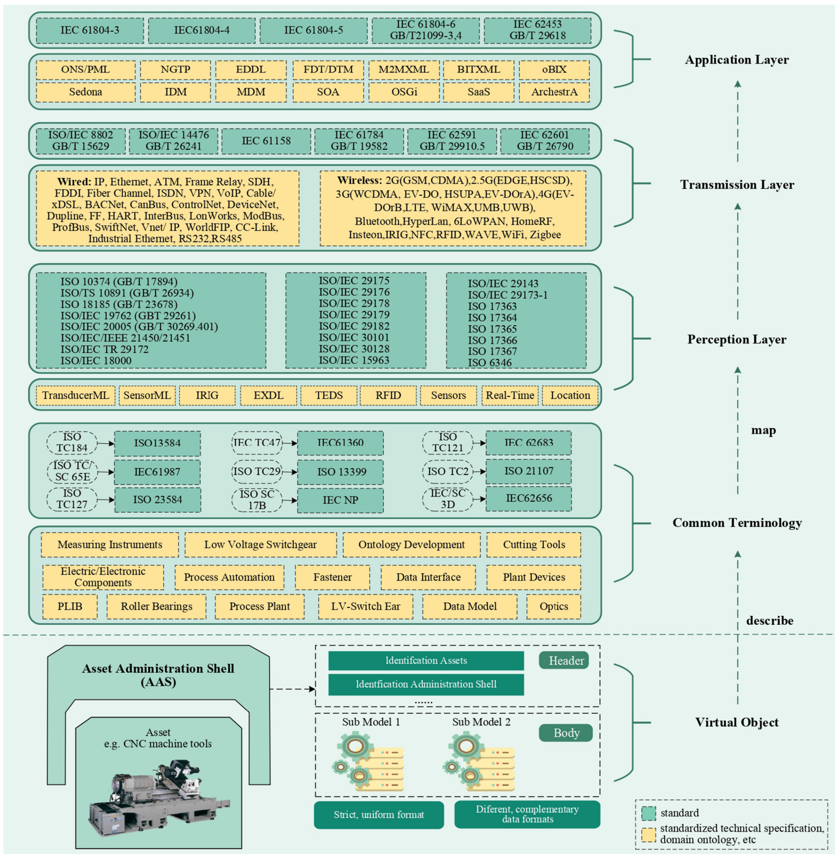

Fig. 3. Standardization in the IIoT.

considering specific domain knowledge, such as communication, configuration, energy efficiency, engineering, identification, life cycle status, status monitoring, and safety and security [254]. These models in different fields need to be formulated and maintained independently by corresponding experts. Each model contains well-structured attributes linked to related data or functions. Such attributes are regarded as a list in the DF body that complies with a standard data format. DF body also has a component manager used to manage and access services in each sub-model.

In addition, virtual interfaces generally provide communication interfaces that allow external components to access data and functions, ensuring continuous access to operational data generated by the physical asset, such as bearing weights and linear axis speeds [256].

2.2 Common Terminology

The basic requirement for representing an object (industrial asset) in the information world is to describe the attributes [267], such as "height," "length," or "width." In the IIoT environment, the attributes of these industrial assets are stored in an ontology and described in general terms. IEC 61360, 61987, and 62683 are the main standards used to describe the attributes of industrial assets. Among them, the IEC61360 issued by IEC SC3D and the ISO 13584 issued by ISO TC184/SC4 jointly developed the PLIB standard, which provides a common method for describing the asset ontology [95]. The description method defines different types of information about the supplier library, such as the resource structure and information model of the asset, the characteristic level and attribute principle of the asset. These regulations represent assets independently of any particular vendor-defined identifier. As shown in Figure 3, the shared dictionary defined by PLIB has been shared as the basis of many asset ontologies, such as measuring instruments, mechanical fasteners, optics and photonics, electronic and electronic components, factory equipment, and environmental declarations [50, 117]. Some industry associations in the private sector also maintain common data dictionaries in a PLIB-compatible form, such as eCl@ss in Europe, JEITA/ECALS in Japan, and EdF/Renault/PSA trio in France. IEC61360-1 details the structure and use of the asset ontology, also known as the "reference data dictionary." IEC61360-2 specifies the dictionary data in more detail, and IEC61360-6 specifies the quality standards for the dictionary's content [224]. Note that the data models defined in IEC61360-2 are also published in ISO13584-42.

There are currently many noteworthy studies related to the PLIB standard. Pethig et al. [175] proposed an information model with an AAS for monitoring the condition of servo motors based on IEC 61360 attributes. The system was implemented with PLCs. By using IEC 61360-compliant attributes, the system could automatically identify the type of servo and configure the appropriate safety thresholds [231]. From an analysis of the potential of ISO 13584, Leukel et al. [136] concluded that PLIB standards were a good choice for product data management.

2.3 Perception Layer

The perception layer involves the definitions and explanations of data sources (sensors, field devices, etc.) for collecting data and knowledge from our physical world [106]. Moreover, the perception layer depends on data acquisition and location-sensing technologies such as sensors, quick response codes, and RFID, namely, **automatic identification and data capture (AIDC)**. AIDC technology solves the issue of real-time access to information and is widely used in the field of IIoT. Therefore, there are many standards for AIDC technology. For example, ISO/IEC 19762:2016 defines and explains the terms in automatic identification technology. TransducerML, SensorML, and TEDS provide definitions for sensors and RFID in the AIDC method and are common technical standards for sensor data exchange. As shown in Figure 3, ISO and IEC provide a series of standards to enable factories to build more reliable and smarter sensor networks [208]. These standards include definitions and requirements for functional layer interfaces and entity-oriented interfaces of smart sensors, as well as methods for cooperative exchange of messages (such as ISO/IEC 29182, ISO/IEC 20005, and ISO/IEC 30101) [96, 100, 101].

In addition, RFID, as a method of automatic identification and data collection, can recognize multiple objects moving at high-speed even in harsh environments without manual intervention [192]. RFID technology has also been used to locate and track humans, even when they are not carrying an electronic device [190, 252]. ISO/IEC 20248 specifies the digital signatures RFID needs to identify and authenticate some data, its source, and the proper method of reading them [97]. ISO/IEC 15459 and ISO/IEC 18000 series provide registration procedures and numbering systems for RFID. In addition, standards such as ISO/IEC 29143 and ISO/IEC 29175 are related to mobile AIDC and RFID and support real-time data capture at the perception layer [98, 99].

2.4 Transmission Layer

The transmission layer is the messenger connecting layers of perception and application through communication networks that are divided into wired and wireless communication networks.

IEC 61158 and IEC 61784 series standards in the wired communication network support fieldbus configuration and configuration files. For example, ISO 15745-3:2003, based on IEC 61158, describes the communication network configuration file and the equipment configuration file of the control system based on IEC 61158; ISO 15745-1:2003 specifies the general elements and rules used to describe the integration model and application interoperability profile and its component profile (process profile, information exchange profile, and resource profile) [230]. The technical standards of wired communication include IP, VPN, **Controller Area Network bus (CANBUS)**, ControlNet, DeviceNet, ModBus, ProfiBus, WorldFIP, Industrial Ethernet, RS232, RS485, and so on. Among them, standards such as ProfiNet, EtherCat, and Powerlink are designed for real-time distributed connections.

IEC 62591 and 62601 standards are used for industrial wireless transmission, which mainly standardizes wireless communications in process automation. The ISO/IEC 8802 series are telecommunications and exchange between information systems, that is, wired or wireless transmission of local and metropolitan area networks. Technical standards for wireless communication include spatially limited networks (e.g., WiFi, Zigbee, Bluetooth, 6LoWPAN, HomeRF, Near Field Communication, and Z-Wave) and 3GPP 2—5th generation mobile communication technology.

2.5 Application Layer

The application layer related to CDA undertakes tasks of calculation, data processing, and mining to realize adaptive control, lean management, and intelligent decision-making. The standards in this layer can be classified as data, software framework, and communication configuration standards.

Data standards describe device information, interface information, and configuration information, such as describing which functional blocks exist in the device type, available parameters, the data types of these parameters, and allowable ranges of these parameters. Data technical standards include **electronic device description language (EDDL)**, Field Device Tool/Device Type Manager, M2MXML, and so on. The EDDL technical standard attempts to build a uniform engineering environment that supports various devices produced by any supplier. As an outside-the-square example of EDDL's usefulness, Banerjee et al. [21] used it to describe databases and files, modeling them as "devices" with their own power to communicate.

Software framework standards describe software component specifications. Technical standards include **Master Data Management (MDM)**, service-oriented architecture, Open Service Gateway Initiative, **software as a service (SaaS)**, and so on.

IEC 62453 series are used to standardize the configuration of the communication interfaces between field devices and control systems, which can configure the communication configuration files of ControlNet, EtherNet/IP, and DeviceNet. IEC 61804 series are responsible for standardizing communication characteristics of smart field instruments and device parameters.

2.6 Summary and Lessons

With the development of IIoT, many industry standards have been developed and iterated to improve uniformity and interoperability for CDA. Equipment, suppliers, factories, and production lines are developed under common and technical standards. This article maps the application areas of IIoT general standards and technical standards to the four-layer structure of terminology-perception-transmission-application. The virtual and digital representation of physical assets with AAS is the basis for data access. Its structure is divided into DF Header and DF Body. Among them, the DF Body contains attributes linked to various data or functions, and such attributes must follow the standard data format. The standard data format is described by common terminology defined by PLIB. The common terminology defined by PLIB has been used in many fields, such as measuring instruments, mechanical fasteners, optics and photonics, electronic and electronic components, factory equipment, and environmental declarations. The standards in the perception layer are related to the construction of data sources (sensors, **Radio Frequency Identification (RFID)** systems, etc.), which are also the underlying standards for IIoT data access. The standards in the transmission layer are used to standardize communication networks, e.g., WirelessHART and WIA-PA, which establish a bridge for its upper and lower layers. The application layer standard is the key to encouraging device interoperability, such as IEC 61804 (EDDL). All the above standards ensure that CDA is performed in a uniform and standardized manner across different layers.

3 DATA ACQUISITION

IIoT captures data from industrial assets such as machines and industrial equipment in assembly lines. In the real world, these industrial assets will have different types, makes, and models of equipment [34]. In addition, the production time and operating environment will be different. Data acquisition refers to obtaining data from the entities (industrial assets) in the network that are used for interconnection, data exchange, and communication through different interfaces and protocols. In the IIoT environment, data comes from different types of devices and represents billions of objects [9]. Newer industrial assets typically have easy-to-obtain mechanisms for data acquisition, while older assets do not. Older industrial assets lack networking and data management capabilities, so obtaining data from them poses even greater challenges.

To obtain data from these industrial assets, we must first understand the functions of the assets and determine the correct parameters that need to be measured according to business needs. After the parameters are determined, the data can be obtained from different types of data sources. Some common data types and sources are:

- Simulate sensor data, such as pressure, temperature, flow, liquid level, and other data.
- Binary sensor data (on/off) from proximity, level, limit, pressure switch, and so on.
- Sensor data through field protocols such as Modbus and CANBUS. Modbus is a commonly used serial communications protocol in industrial applications. CANBUS is a standard protocol that is widely used in the automotive and heavy equipment industry.
- Wireless sensor data from sensors through Bluetooth and wireless HART.
- Data from PLC, SCADA systems.
- Data from equipment controllers such as engine control units and dedicated controllers used for operating equipment.
- Data from 3rd party cloud platforms through application programming interfaces.
- Data from the web, such as weather and air quality.
- Data from enterprise systems (asset master data, maintenance data, spares data). In an industrial environment, enterprise systems include enterprise resource planning systems, maintenance management systems, spare parts management systems, and so on.

3.1 Data Types in IIoT

The acquired data can be roughly divided into the following types:

- Direct Sensor Data Acquisition: It is about acquiring data from standard sensors such as pressure, temperature, flow, proximity, level, and so on. In particular, sensor data and wireless sensor data acquired through the field protocols also belong to it. Besides, derived measurements are done if direct measurements are not possible due to physical, technological, or cost constraints. This is done by measuring a suitable proxy parameter and deriving the target parameter from this proxy parameter.
- Data from Industrial Control Systems: PLC, DCS, programmable automation controllers, computerized numerical controllers, motion controllers, and SCADA systems are some of the most commonly used **Industrial Control Systems (ICSs)**. OPC UA, EthernetIP, Profinet, Modbus, and serial interfaces are some common interfaces for obtaining data from ICSs.

Although interfaces such as OPC UA, EthernetIP, and Profinet are used in the newer versions of ICSs, older legacy systems still use serial interfaces and Modbus.

— Environment data: It is mainly produced by enterprise systems and includes natural and human environment data. Natural environment data is about the description of the environment and their combination and processing data, such as machine ID and machine type. Human environment data involves people in managing production, such as order information.

The following will introduce three methods for solving the above-mentioned data collection problems: plug-and-play architecture, reconfigurable scheme, and hardware expansion.

3.2 Plug-and-play Architecture

Plug-and-play architecture is an important solution to the problem of direct sensor data acquisition. Especially the problem of wireless sensor data collection. To support plug-and-play architecture, the industry has proposed a unified family of standards—the **Institute of Electrical and Electronics Engineers (IEEE)** 1451 [211]. IEEE 1451 is a set of smart sensor interface standards for connecting sensors, actuators, microprocessors, and/or instrument systems in a control or field network through a standard set of communication interfaces [180, 211]. Many researchers have used this system in their designs for applications such as environment monitoring systems [128, 130, 235], precision agriculture [62, 232], and intelligent vehicles [127, 135].

Q. Chi et al. [42] adopted a **complex programmable logic device (CPLD)** as the main controller of an IoT environment and proposed a new method for designing reconfigurable smart sensor interfaces for industrial **wireless sensor networks (WSNs)**. The device combines the latest CPLD programmable technology with IEEE 1451.2 smart sensor specification standards. Bordel et al. [26] realized a certain degree of multi-source sensor data collection based on the IEEE 1451 standards by designing a configurable multi-sensor acquisition interface. IEEE 1451 has been applied as a standard interface to an environmental monitoring system [129], a green building [154], and digital twins [210]. Both Cherian et al. [41] and Guevara et al. [76] used standardized **transducer electronic data sheets (TEDS)** specified in terms of IEEE 1451.2 in their applications to do with data perception. The definition of various transducers through these TEDS is a key element of the IEEE 1451 standards. TEDS contains the information required to act as an interface between a measuring instrument, like a sensor, and the control system. Sometimes, they are embedded on a storage device connected to the sensor, but they can also reside in embedded memory on the sensor itself, such as on an EEPROM [41].

The **Open Geospatial Consortium (OGC)** develops the Sensor Web Enablement standards for building a sensor web to discover and use sensors flexibly [27]. Based on these standards, Bröring et al. [29] developed a sensor plug-and-play infrastructure to integrate sensors on-the-fly with minimal human intervention.

Therefore, although some researchers draw on IEEE 1451.4 [140], others have turned to the OGC [27] and still others have turned to TEDS [41, 42, 76], especially analog transducer users in this latter case. These plug-and-play standard solutions can help a lot to solve the direct sensor data acquisition problem. However, the plug-and-play architecture has a limited scope of application.

3.3 Reconfigurable Scheme

The reconfigurable scheme is mainly used to solve the collection problem of data from equipment and data from ICSs. It also supports sensor data from field protocols. The reconfigurable scheme utilizes the system characteristics of programmable to realize a universal data acquisition system. For a universal data acquisition system, in the face of many data items that are very different from each other and tightly coupled to a specific system, a manually coded protocol parser is difficult to reuse. Liu et al. [141] provided a framework for automatic online application protocol field extraction that can be parsed from context-sensitive application protocols and can also generate optimizations based on simple extraction specifications. Adesina et al. [7] and Graham et al. [73] selected a deterministic finite-state automaton as the general protocol analysis engine.

Some single-chip computers are also used in universal acquisition systems due to their programmability, such as **field programmable gate array (FPGA)**, Raspberry Pi, and PLC. Different ways of using FPGA hardware to collect data have been explored in many studies [22, 67, 82, 126, 179]. One of the most difficult problems to address has been overcoming the fact that every device manufacturer designs and uses its own bespoke protocol with no concern for compatibility or communication with out-of-house units. Bao et al.'s [22] solution to this issue is a reconfigurable data acquisition system for industrial sensors that uses an FPGA as the core controller. Wu et al. [239] also proposed a dynamic data access system but with embedded PLCs that have flexible hardware and software structures [240, 241].

The universal data acquisition system makes use of configurability to realize the parsing of various field protocols. However, programmable devices still have resource and performance limitations.

3.4 Hardware Expansion

The last of the three approaches is to extend a device's physical interface with a specific fieldbus. In fact, most manufacturers offer specialized hardware extension schemes. Schneider's TM3 Bus Coupler series serves as an instance of hardware extension, aiming to improve the performance of the Modicon M221, M241, and M251 logic controllers. Alternatively, the Schneider STBXCA100X series is a Modicon STB distributed I/O solution, divided into different products according to length. Siemens provides proprietary modules SIMATIC IoT2000 series to extend their product lines to the IIoT environments. Omron's CP1W units can expand I/O units, LCs, and any other devices via Ethernet [170, 205]. Omron introduces EtherCAT as the machine control network. EtherCAT can adopt linear, star, and ring topologies, and the possible expansion on each node helps to achieve the maximum network flexibility of machine design [170]. However, hardware expansion cannot be used as a general solution for IIoT because of its high price and physical inflexibility.

3.5 Summary and Lessons

As shown in Table 3, the methods of collecting data are mainly divided into plug-and-play architecture, reconfigurable scheme, and hardware expansion. Among them, plug-and-play architecture is a way to address the issue of direct sensor data acquisition, especially the wireless sensor data collection problem. Its advantage is that IEEE 1451 standardizes the plug-and-play interfaces for data acquisition. However, the disadvantage is that it is not applicable to data sources such as ICSs. Reconfigurable schemes are mainly used to solve the collection problem of data from equipment and data from ICSs. In particular, it also supports sensor data from field protocols in direct sensor data acquisition. The programmability of some single-board computers or microcontrollers (e.g., PLC, FPGA, and Raspberry Pi) is used to implement a reconfigurable system, which can be used as a general data acquisition system. However, programmable devices still have resource and performance limitations. Hardware extension is a physical extension method proposed by many manufacturers, which can be applied to collect data from various data sources. However, hardware expansion has the disadvantages of high price and insufficient flexibility.

4 DATA FUSION

Data fusion refers to the techniques and strategies for overcoming heterogeneity in data formats and sources. The heterogeneity in data can take two forms: the data format, such as text data,

Table 3. Data Acquisition

Solution	Author	Application	Types: S	E	I	Description	Limitation
Plug-and-play Architecture	Kumar et al. [129]	Environment monitoring system	✓	✓	-	IEEE 1451 standardizes the plug-and-play interfaces for data acquisition.	Configurable interfaces that comply with IEEE 1451 only work with some types of sensors.
	Wei et al. [235]	Precision agriculture	✓	–	–		
	Lee et al. [135]	Intelligent vehicles	✓	–	✓		
	Maiti et al. [154]	Green building	✓	–	–		
	Chi et al. [42]	Water environment monitoring	✓	✓	–		
	Song et al. [210]	Digital twins	✓	–	–		
Reconfigurable Scheme	Garcia et al. [67]	IoT environment	✓	–	✓	Programmable devices, such as FPGAs and PLCs, are used in universal access systems because of their programmability.	Programmable devices have resource and performance limitations.
	Wu et al. [240]	Automatic winding machine	✓	–	✓		
	Bao et al. [22]	Industrial production control	✓	✓	✓		
	Portilla et al. [179]	Environment monitoring system	✓	✓	–		
	Hinkelmann [82]	IoT environment	✓	–	✓		
	Sundaramurthy [216]	IoT environment	✓	✓	–		
	Shashikant [201]	IoT environment	✓	–	✓		
	Ursutiu [225]	IoT environment	✓	✓	–		
	Wan [228]	Drug Packing	✓	–	✓		
	Krasteva et al. [126]	Industrial IoT environment	✓	✓	✓		
Hardware Expansion	Schneider [195]	Schneider TM3	✓	✓	✓	Extending physical interfaces.	Hardware expansion is expensive and not flexible enough.
	Siemens [205]	SIMATIC IoT2000	✓	✓	✓		
	OMRON [170]	Omron's CP1W	✓	✓	✓		
	Siemens [205]	Schneider TCSXCNNXNX100	✓	✓	✓		

S = Direct Sensor Data Acquisition, E = Data from Equipment, I = Data from ICSs.

relational data, image data, and audio and video data [262]; or its structure, namely, structured, semi-structured, or unstructured data [119]. In industrial environments, and in big data generally, most data are unstructured. They have vagueness, which means the information suggests little about what it conveys [262]. To illustrate, the word "wind" on its own is vague. It could be a noun or a verb; it could relate to air, a path, a piece of string, a stomach pain. Without context or further information, we cannot know what this word is intended to convey. For our purposes, we define heterogeneous data as data from different sensor sources, with different collection periods or types but that relate to each other.

Data fusion technology structures and integrates heterogeneous data from different sources to improve its comprehensibility, accessibility, and extensibility. This is an active area of research for both academia and industry, with the main spheres of inquiry being compressed storage, fault prediction, location, and status assessment. Descriptions and some representative examples of these research themes follow.

Storage compression refers to reducing the size of the data and tries not to lose useful information. For example, in pursuit of more efficient storage solutions, Park et al. [174] applied machine learning to process data into a vector representation of only a single formula. Cheng et al. [39] proposed a lossy fast lifting wavelet transform to compress distributed sensing data for large-scale WSNs. This strategy largely reduced the amount of data that needed to be transmitted.

Fault prediction is used to detect faults by specific sensors. Here, Zhou et al. [269] devised a multi-sensor global feature extraction method for monitoring the condition of various tools in a milling process based on kernel extreme learning machine and a genetic algorithm, while Zhou et al. [268] fused a multimodal feature data to extract abnormal frequency features for rotating machinery diagnosis.

Location studies focus on identifying the precise location of objects. Zhang et al. [261] developed a localized fingerprint technology based on heterogeneous feature fusion and the time-of-arrival feature to enhance the precision and robustness of indoor positioning. Alatis et al. [12] proposed a method of estimating the position and orientation of mobile robots based on the fusion of inertial sensor data and vision.

Status assessment is to give a comprehensive evaluation from many indicators. In this stream, Ha et al. [78] proposed an air quality assessment system that merges many indoor air quality index, such as CO, $PM_{2.5}$, and SO_2. Sun et al. [214] proposed an intelligent data fusion framework

Table 4. Surveys in Data Fusion Application and Framework

C1	Task/Type	Author and one-sentence summary	Advantages						
			P	E	S	Y	L	M	C
Application	Compression	Park et al. [174]: A machine learning-based method to process industrial data into a representative vector represented only by one formula for reducing storage consumption	–	–	–	✓	–	✓	–
		Bardwaj et al. [23]: A method of compressing data at each local sensor to improve the accuracy of fusion estimation	–	–	–	✓	–	✓	–
		Cheng et al. [39]: A distributed fast lifting wavelet based on the lossy data compression algorithm	–	–	–	✓	–	✓	–
	Fault Prediction	Zhou et al. [269]: A multisensor global feature extraction method	✓	–	–	✓	–	–	–
		Zhou et al. [268]: A multimodal feature data fusion method for rotating machinery diagnosis	✓	–	–	✓	–	–	–
	Localization	Zhang et al. [261]: A fingerprint localization technology based on heterogeneous feature fusion and the time of arrival feature	✓	–	–	✓	–	–	–
		Alatis et al. [12]: Fusion of inertial sensors data and vision	✓	–	–	–	–	–	–
	Status Assessment	Ha et al. [78]: An air quality management system	✓	✓	–	✓	–	–	–
		Sun et al. [214]: An intelligent data fusion framework	✓	✓	–	–	–	–	–
Framework	Edge-based	Izadi et al. [102]: A fusion approach based on fuzzy logic for wireless sensor networks	–	✓	✓	–	–	✓	–
		Yang et al. [251]: A temporal data fusion method based on Gaussian processes	–	✓	✓	–	✓	–	✓
		Larras et al. [132]: A sensor-level clique-based neural network that serves as a generic classifier	–	–	✓	–	✓	–	–
		Qian et al. [181]: Real-time fault diagnosis on edge nodes	–	–	–	–	✓	✓	✓
	Cloud-based	Mu et al. [202]: An intelligent transportation control system implemented as a cyber-physical cloud application	–	–	–	✓	–	–	–
		Hao et al. [244]: An information fusion system based on cloud computing with greater resilience and a capacity to process data	–	–	✓	–	–	–	✓
		Liu et al. [142]: An architecture for teaching cloud-based robotic systems how to navigate	✓	–	–	–	–	–	✓
		Sirisakdiwan et al. [207]: A Spark streaming framework to support multi-heterogeneous streams being deployed in a single Spark application	–	–	✓	✓	–	✓	–
		Stojanovic et al. [212]: A software platform for remote health	✓	–	–	–	–	–	✓
	Collaborative	Asif et al. [16]: A five-layered IoHT framework for time-critical data processing	–	–	–	✓	–	–	✓
		Wang et al. [233]: A fog-based environmental monitoring framework for fusing multi-source heterogeneous data locally	✓	–	✓	–	✓	–	✓
		Arooj et al. [15]: A five-layered collaborative multimodal architecture with data fusion functions based on context descriptors, tensor decomposition, and semantic approaches	–	✓	–	✓	–	✓	✓
		Valente et al. [226]: A container-based microservice for fusing multi-sensor data	–	–	–	–	✓	–	–

C1 = Category, P = Precision, E = Energy, S = Safety, Y = Yield (efficiency), L = Latency, M = Memory (Storage), C = Computing.

to compute the health status of a bridge according to an index. The framework comprises a hybrid neural network built on adaptive resonant theory and an adaptive fuzzy inference system.

In real-world scenarios, deploying an efficient data fusion framework not only requires an appropriate algorithm but also a systematic framework to support data collection, analyses, and decision intelligence. Table 4 lists several surveys on data fusion applications and frameworks. In this section, we provide examples of contemporary data fusion in terms of three frameworks: edge-based, cloud-based, and collaborative frameworks. Note that collaborative frameworks focus more on cooperatively processing data compared with edge-based and cloud-based ones.

4.1 Edge-based Frameworks

Izadi et al. [102] proposed a fusion approach based on fuzzy logic, which is applied in a WSN attempting to improve the QoS and reduce the energy consumption of the edge WSN. The strategy is to reduce the amount of data transmitted through fuzzy rules that distribute and collect only true values. This eliminates redundant data, which consequently reduces energy consumption and increases the lifespan of the WSN. Yang et al. [251] argue that transferring data from a sensor node to the cloud server without prior processing raises several concerns, including high latency, privacy leaks, and excessive bandwidth consumption. To address these issues, they devised a method of fusing temporal data based on a Gaussian process. Essentially, the methods make predictions from time-series data streams at the edge device so only the temporal features of the stream need to be transferred between the cloud and edge. Larras et al. [132] suggest shifting the small-scale classification and feature extraction parts of the processing to the sensors using **clique-based neural networks (CBNNs)** as a generic classifier. CBNNs allow the method to be divided into clusters and dispatched to different nodes of a WSN, which would help to reduce communications overheads as well as ensure privacy. Qian et al.'s [181] solution for a system of rotating machines involves real-time fault diagnosis and dynamic controllers deployed on edge computing nodes. Most of the data are computed on **electronic communications networks (ECNs)**, where the controllers can modify the operational parameters if needed. Only important data are sent to the server for further analysis, which reduces storage and computing requirements on the cloud. The features are extracted via a fast Fourier transform and fused with a classical backpropagation neural network.

4.2 Cloud-based Frameworks

Due to the constrained resources of edge devices, data fusion methods performed at the edge tend to be fixed, uncomplex, and non-interactive. Hence, most complex data fusion tasks with changing demands are either performed by a central server or designed for centralized computing architectures.

Mu et al. [202] designed an intelligent transportation cyber-physical cloud control system, where dynamic information from heterogeneous sensors is sent to a central control platform. The data processing and analysis are fused within an intelligent transportation network, where the prediction results and control schemes are generated simultaneously and sent to the terminal systems to support unified monitoring, management, decision-making, and control services. Hao et al. [244] claim that data throughput and computing power in traditional information fusion systems cannot meet the demands of future competition. By traditional, they are referring to systems based on a single platform with a small number of sensors. As such, this article puts forward a design and a highly detailed workflow plan for an information fusion system based on cloud computing with more resilience and greater capacity for data processing. Liu et al. [142] present an architecture for teaching cloud-based robotic systems how to navigate. The learning scheme follows **lifelong federated reinforcement learning (LFRL)** to support reducing training time without sacrificing accuracy. With LFRL, robots in different environments can fuse and share models to help each other learn more quickly.

To overcome issues with high latency, many researchers have turned to cluster frameworks, such as Spark [258] and Hadoop [204]. These packages offer as near to real-time cloud computing as is currently possible and are, therefore, very popular. Sirisakdiwan et al. [207] claim that, at present, there are many problems with multi-stream applications, including difficulties with deployment and monitoring, redundant coding, and inefficient job queueing. To overcome these issues, they propose a Spark streaming framework that supports the deployment of multiple heterogeneous streams in a single Spark application. They use the case of anomaly detection in a network to evaluate the effects of fusing all these streams. Stojanovic et al. [212] developed a software platform, RemoteHealth, to provide support for the collection, fusion, processing, and analysis of large-scale data. Their application also involved human activity recognition on a dedicated server within the RemoteHealth platform, all based on Apache Spark.

4.3 Collaborative Frameworks

Although the Spark framework allows data fusion with real-time streams, many researchers implement these tasks on a distributed system for better performance. For example, Asif et al. [16] presented a five-layered **Internet of Health Things (IoHT)** framework, consisting of mist, fog, and cloud layers. The mist computing layer at the extreme edge of the network performs rapid rule-based preprocessing of the sensor data with basic tasks, such as data aggregation, fusion, or filtering, which is key to ensuring time-critical data processing.

Wang et al. [233] claim that real-time monitoring is barely possible with a centralized cloud solution, because the data collected by a single edge node is insufficient to accurately describe the state of an environment. Their solution is a fog-based environmental monitoring framework, which can fuse multi-source heterogeneous data locally. A cloud data training system is responsible for updating the fusing models in different edge nodes. However, rather than sending an entire model, the edge nodes only transmit the model parameters as an update, not the raw data.

Arooj et al. [15] proposed a five-layer collaborative multimodal architecture. In particular, the third layer is dedicated to knowledge discovery with the collected data. They consider that an increased number of nodes results in higher communication overhead, so having one central server process all the data is neither feasible nor efficient. Their data fusion solutions are based on context descriptors, tensor decomposition, and semantic approaches. Valente et al. [226] implemented a container-based microservice to fuse multi-sensor data, which is configured as fog architecture and built from open-source IoT middleware. With this strategy, the scale of the system can adjust automatically when excessive data is input from the edge nodes.

4.4 Summary and Lessons

Collected data from the IIoT environment is massive and heterogeneous, which is challenging the data fusion for further processing. In terms of the applications, the fusion can be performed in three frameworks: edge-based, cloud-based, and collaborative. The edge-based framework is suitable for applications with a small amount of data but real-time requirements. The cloud-based framework can process complex and high-computational data fusion but loss the real-time capability. Collaborative frameworks leverage both advantages of cloud and edge features to fuse data. As the last stage of data access, data fusion prepares well-structured data for further applications, and as a whole process, data acquisition and fusion or even preprocessing such as data cleaning can be done in one framework in a flexible way. For instance, the data stream cleaning system proposed by Sun et al. [215] executes the data collection, fusion, and cleaning sequentially and scalably, where the programs of the above three stages can be updated according to requirements. However, their scalable mechanism is not general. This is also the reason we investigate virtualization technology and provide a discussion in the following section.

5 SCALABLE ARCHITECTURE

Factories add new production lines, and transportation networks add new hubs and vehicles to their fleets. So, too, IIoT architectures need to be able to scale to cater to this growth. Moreover, every new entity in the system is likely to come with its own data processing logic, and this growth in heterogeneity also needs to be accommodated. In this section, we review progress toward scalable architectures to support these requirements. Beginning at the dawn of networking with the first central server, we review the many and varied rules, methods, functions, and implementations of computer systems up to and including the virtual technology of today.

5.1 Evolution of Computing Architectures

In the past few decades, various types of computing paradigms and architectures have emerged to suit almost every purpose under the sun.

Supercomputing is the original purpose that originated from the invention of the first computer to the first supercomputer UNIVAC LARC in 1960. However, as impressive as these large proprietary mainframes with high-performance computing capabilities were, they could still only perform one task at a time. To perform multiple tasks, multiple systems were needed, each running in parallel. Hence, the 1970s saw the advent of cluster computing, open massively parallel processing, and symmetric multiprocessing. In the 1980s, grid computing emerged [80], and this concept dominated academia right through until the early 1990s [64]. It was in this decade that architectures truly began to diverge, especially in terms of data storage. Some frameworks focused on centralization, others on decentralization, but, over the next decade, we would witness the evolution of commodity clusters, peer-to-peer, web services, virtual clusters, HPC systems, laaS, PaaS, and SaaS. All these ideas would culminate in 2006 with the popular term "cloud computing" because of Amazon's Elastic Compute Cloud product. In its initial conception, cloud computing was typically implemented as a joint hardware and software platform, designed to let businesses share resources and make high-priced computing services affordable to SMEs [113].

The next few years would see a tumult of innovation. The Internet of Things was born during this time as an interconnected system of edge devices with high data processing requirements [167], as were fog computing, edge computing, and edge-cloud collaborations.

Fog computing, which appeared in 2011, uses terminal devices, i.e., fog nodes, to bear most of a system's storage and calculation requirements and the internet as the communications medium [17]. Fog platforms include fog nodes and cloud servers. The fog node has a certain amount of calculation ability, which provides real-time and **quality of service (QoS)** services near the IIoT environment, while the cloud servers support the distributed fog nodes with elastic and huge computing power [25].

Edge computing surfaced next. Edge computing can be traced back to the 1990s, but it has been widely used recently due to the dramatic performance improvement of edge hardware. As a distributed computing paradigm, the edge devices collect the data and perform some or all of the calculations required before transmitting the results through the network [260]. Because much of the computing is done at the edge, these systems typically have lower latency, lower transmission costs, better response times, and better QoS than conventional cloud systems. For this reason, they are a popular choice for applications with large-scale mobile data [37, 193]. Time-sensitive applications can also benefit from edge computing, particularly with a highly reliable and available internet connection [209]. The major shortcoming of edge computing is that the edge devices are generally resource-constrained, which means their computing power is limited to simple, non-intensive tasks [203].

The current trend for integrating time-sensitive applications within the resource limitations of edge devices is an edge-cloud collaboration. These frameworks combine the real-time performance of edge computing with the powerful functions of cloud computing, giving play to the advantages of both paradigms.

Currently, research interest in edge-cloud collaboration is high, which is resulting in very rapid developments in this area. However, in general, edge-cloud computing frameworks comprise the edge layer, the communications layer, and the cloud layer [35, 45, 52, 159, 182, 208, 219, 229, 248]. Some architectures include an additional fog layer between the edge and the cloud. Gateways aggregate and transfer data from the physical devices in the edge layer, through the communications layer, to the cloud layer at the top. The connections between layers can be wired or wireless. The edge layer might include devices, such as sensors, controllers, and actuators; systems, such as **supervisory control and data acquisition systems (SCADAs)**, DCS, and **manufacturing execution systems (MES)** [59, 159]. Example protocols used for the communication layer include **Internet protocol version 6 over a low-power wireless personal area network (6LoWPAN)**, **message queue telemetry transport (MQTT)**, **extensible messaging and presence protocol (XMPP), advanced message queuing protocol (AMQP), data distribution service (DDS), constrained application protocol (CoAP)**, and **open platform communications unified architecture (OPC UA)** [59, 196]. Also, emerging technologies such as network virtualization, **software-defined networking (SDN)**, and **fifth generation mobile networks (5G)** are playing important roles in this layer.

Table 5. Surveys in Virtualization Technology

C1	Methodologies	Advantages	Disadvantages	Contributions
Full	J. Hwang et al. [94]: An ARN CPU architecture for virtualizing full systems and a prototype implementation of a hypervisor, called Xen.	• High reliability • High performance • Moderate cost	• Low scalability	• Advantages • Challenges • Design principles • Current mainstream product • Application
	Nishikiori, Masaaki [169]: A combination of Fujitsu's high-reliability servers, middleware, and vSphere 4 VMware virtual infrastructure for optimizing data center operations.	• High scalability • High reliability	• High execution time	
Para-Virtualization	Jun Zhang et al. [114]: An embedded architecture based on a kernel virtual machine that combines VxWorks and Linux for real-time virtualization.	• Low interrupt response latency • Strong isolation	• API support	• Advantages • Challenges • Comparison with full virtualization • Current mainstream product • Application
	Bugnion et al. [30]: An x86 architecture for virtualization with VMware Workstations.	• High compatibility • Simple user experience • Low overall overheads	• API support	
Container Technology	A. Ahmed et al. [8]: Comprehensive improvements in Docker by an equential image downloading, a multi-threaded decompression, and I/O pipeline for small computers, such as Raspberry PIs.	• Low development time • Good scalability • Good isolation	• Limited by network bandwidth, CPU, or disk I/O	• Structure • Major components • Advantages • Challenges • Current mainstream product • Application
	J. Tang et al. [221]: It makes the abstraction and management of the execution environment in the granularity of containers on edge.	• Lightweight • High feasibility	• Limited in the large-scale environments	
	Yzhou Huang et al. [92]: A computing architecture based on Docker and Kubernetes for machine learning.	• Rich resources • Stable network	• High resource consumption	
	Jaiswal et al. [103]: A patient data collection framework based on the container intelligent medical system	• Good scalability • Lightweight	• Lack of evaluation of user constraints	
Container Orchestration	Kaur et al. [120]: A competent controller for container management on edge-cloud nodes that considers Co2 emissions, carbon footprints, and energy consumption.	• Good green energy utilization • High performance • High flexibility	• Ignores network aspects of QoS, e.g., latency	• Advantages • Challenges • Examples of platforms products • Current mainstream product • Application
	Dupont et al. [55]: A platform can perform to distribute functions horizontally and vertically.	• Good remote diagnosis • Roaming IoT function	• High resource consumption	
	Chen et al. [36]: An architecture based on Kubernetes-enabled NFV and management and orchestration for implementing a scalable IoT/M2M system in the OpenStack cloud.	• Good scalability • Good management and orchestration	• High resource consumption	

5.2 Virtualization Technology

Virtualization is the key to achieving dynamic and scalable services. Virtualization technology is a kind of resource management technology that abstracts the physical resources of a computer (e.g., networks, memory, CPUs, storage, I/O equipment) in software manner, overcoming the limitation caused by the indivisibility of physical resources. Virtualization technology enables a single physical computer to be transformed into several logical computers, realizing dynamic scheduling and multi-scale sharing of physical resources, which in turn improves resource utilization. In its early form, virtualization simply meant dividing a system's resources and purposing them to different applications. Today, virtualization can mean anything using software to mimic hardware to support larger and more complex applications. A list of studies on virtualization is provided in Table 5.

Virtualization technology mainly includes virtual machines, containers, and container orchestration engines. The types fall into two broad categories: full virtualization and para-virtualization [14]. Full virtualization manages virtual machines through an additional intermediate software layer between the virtual machine and the hardware called a hypervisor [188]. With full virtualization, the underlying hardware is fully simulated, which means native applications and software can usually run transparently and seamlessly. However, because the hypervisor uses resources, which adds overhead to the system, performance is not as good as if one were using the actual machine.

Para-virtualization essentially adds some interfaces so users can run software and apps designed for another system on their own host environment. It gives the appearance of operating in a particular environment, but the artifice is superficial. However, because para-virtualization modifies the operating system kernel to use the application programming interface provided by the virtualization layer hypervisor instead of nonvirtualizable instructions, it eliminates the need for a **virtual machine monitor (VMM)** translation process. Therefore, para-virtualization generally performs better than full virtualization.

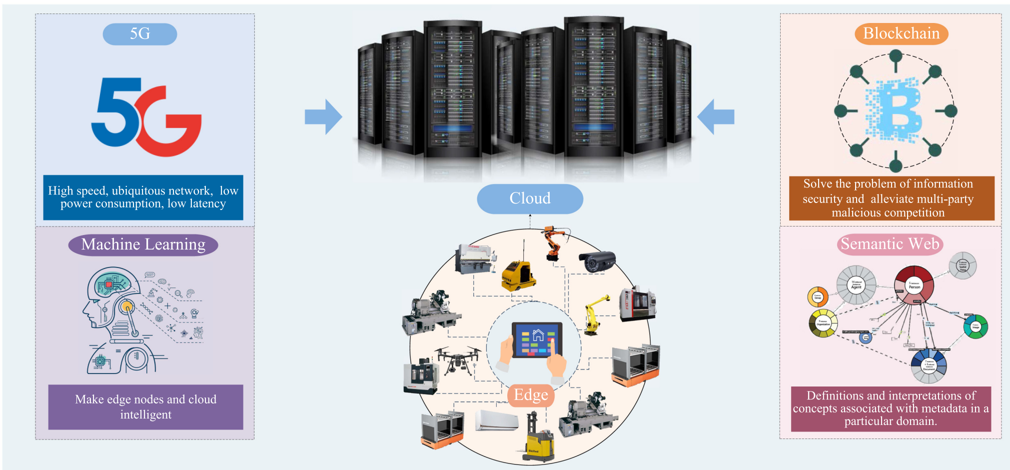

Fig. 4. Emerging fields with relevance to CDA.

5.2.1 Virtual Machines. The most widely used virtualization technology is the virtual machine. Virtual machines come in two main flavors: one runs on bare machines, e.g., Xen, VMware vSphere, and KVM; the other runs on an operating system, e.g., VMware Workstation. Both are essentially a copy of an application running in parallel on the same server, which is an efficient way of using hardware resources. It also makes many application services more feasible and/or more robust. For example, virtual environments are less prone to interruptions of service, and, when services are interrupted, they tend to recover more quickly. Upgrading from old to new software versions is easier—one simply installs the new version on a new virtual machine and flips over to the new machine when the time is right. Also, the isolation inherent to virtual machines means different applications are less likely to interfere with each other. Further, virtual machines decouple an application from an actual physical server, which means the server running the service can be changed dynamically to suit prevailing circumstances. The process of changing servers is called fast migration [145]. However, perhaps the greatest benefit of virtual machines is that users can use as many or as few resources as needed to suit their current workload requirements [178]. Scaling programs by adding and deleting resources is part of the application execution process. Today, there is even virtualization technology for deploying and managing the cloud platforms themselves. OpenStack is a pre-eminent example [32].

5.2.2 Containers. In addition to virtual machines, containers are another virtualization technology that has emerged in recent years [53]. They are like virtual machines but structurally different. As shown in Figure 4, virtual machines and containers both comprise basic hardware facilities, i.e., a physical **host machine running an operating system (HostOS)**, binary files, development dependent libraries, and an application service layer. But a virtual machine includes a virtual machine operating system (GuestOS) and operates on top of the VMM/hypervisor, while a container runs upon the container engine, which in turn runs on the operating system. In other words, a virtual machine must simulate a complete operating system, while a container is just an instance share of the native operating system running on the physical machine.

Up to now, many container technologies have emerged, which can be mainly divided into four categories according to their functions: container runtime (such as CRI-O and Containerd) [60], container engine (e.g., Rkt, OpenVZ, LXC) [115], application container engine (like Docker) [246], and container orchestration engine (such as Kubernetes and Apache Mesos) [107].

(1) The container runtime implements the **Container Runtime Interface (CRI)** specification, which is used to manage the deployment and execution of containers.
(2) The container engine provides a lightweight virtual runtime environment at the **operating system (OS)** level, using the same kernel mechanism as the OS. This technology allows for the virtualization of a physical machine into multiple OS instances.
(3) The application container engine is an application encapsulation technology that does not have its own kernel or virtual hardware. It only encapsulates the application and its dependent environment as an independent object, achieving isolation of application processes. The application in it runs directly on the host kernel.
(4) The container orchestration engine is used for the automatic deployment, scaling, and load balancing of container-based applications.

Containers have three main advantages:

(1) They are lightweight. Because the containers themselves do not have a kernel, they consume fewer resources than a virtual machine. Also, this means container startup is usually a second-level process, which is more suitable for the rapid startup requirements of most industrial environments. Plus, it means application services can be deployed and migrated quickly. Salonidis et al. [153] introduced dynamic service migration based on containers in a mobile edge-cloud framework. Li et al. [150] introduced a rapid migration architecture for edge services based on containers. Kakakhel et al. [116] defined the different properties of containers and discussed how to dynamically migrate stateful applications through containers to extend fog and mobile edge computing services.
(2) They have good scalability. Different types of devices can automatically pull updated versions of an image file to update a container through a container engine, destroying the old version of the container to release resources. Darrous et al. [46] proposed two container image placement algorithms, **k-Center-Based Placement (KCBP)** and **KCBP-Without-Conflict (KCBP-WC)**, to reduce the maximum retrieval time of a container image across edge servers. KCBP and KCBP-WC are optimized based on their K-centers.
(3) They have good isolation. Isolating applications with containers, both physically and in terms of system resources, is largely a function of Cgroups and namespace technology. This is a highly cohesive system that can rapidly separate services, which perfectly suits microservices architecture. Ren et al. [139] provide an excellent explanation of the isolation performance of containers.

Unlike cloud computing, the edge nodes in edge computing have less CPU power, memory, storage, and networking capabilities, so any technology installed on an edge device to make it more scalable must be lightweight. For this reason, container technology is suited to edge applications and, as such, an abundance of solutions have already been developed. For instance, Ahmed et al. [8] proposed three different Docker optimization methods to improve the efficiency of hardware resource use in a fog architecture that involves very small computing nodes, such as Raspberry Pi. The three optimizations, sequential image downloading, multi-threaded decompression, and input/output pipelining, each combine to reduce a server's deployment time by approximately 30%. Huang et al. [91] introduced a technology that combines containers and edge computing around parked vehicles to improve resource allocation in transport systems. Tang et al. [221] proposed a container-based edge unloading framework for a self-driving application. The framework includes an offload decision module, an offload scheduler module, and edge offload middleware based on lightweight virtualization. Jaiswal et al. [103] provided a collection framework for patient data in an intelligent medical system based on Raspberry Pi-configured containers. Mendki et al. [158] applied Docker to the analysis of IoT video at the edge in a study on the feasibility of analyzing surveillance video in real-time with a deep learning framework, again, based on Raspberry Pi. Their performance metric was the cost of containerization. Hsieh et al. [88] demonstrate various application scenarios in an edge computing platform with the support of containers, including service deployment for smart cities such as air quality monitoring and voice and image recognition/classification. They tested deployment speeds, QoS levels, and the response times of some event-driven mechanisms.

Navarro et al. [19] introduced the idea of using containers to deploy services from a local server as a micro edge-cloud collaboration. They show how a container approach can help third parties create and share customized services at the edge of a network to better meet specific local needs and constraints. They also illustrate that containers in edge-cloud environments are particularly suited to PaaS systems due to their lightweight size and flexibility. Lee et al. [173] demonstrate that container and clustering technology can promote the applicability of applications on the distributed multi-cloud platform collaborating with a series of network nodes.

In fact, today, containers have become so interwoven with distributed computing that many commercial edge devices are either containerized or container-ready, e.g., Azure IoT Edge [161], Balena [20], and Amazon's AWS IoT Greengrass [13].

Azure IoT Edge was announced as an open-source product by Microsoft in 2018. It comprises certified Edge hardware, modules, a runtime, and a cloud interface. An edge module is an execution unit, implemented in a Docker-compatible container, that runs business logic at the edge. The runtime allows for the use of custom and cloud logic on edge devices. It sits on the edge device and performs management and communication operations. The cloud interface remotely monitors and manages the edge devices.

Balena consists of BalenaCloud, BalenaOS, and BalenaEngine. BalenaCloud is Balena's fully managed cloud service. In addition to a server that can automatically package code into containers, it includes device and client software. BalenaOS is a simple Linux operating system suitable for edge devices that comes with the BalenaEngine pre-installed. The whole suite forms a lightweight, Docker-compatible container engine.

AWS IoT Greengrass enables customers to build edge devices and applications. With this kit, devices can securely connect to a local network and interactive information without requiring to communicate with the cloud.

5.2.3 Container Orchestration. Container orchestration allows containers to run as a cluster in distributed environments. An orchestration engine usually includes a container management mechanism, a scheduling mechanism, cluster definitions, and a service discovery module. Through these components, the engine organically combines containers into microservice applications to fulfill business needs. Containers can even be expanded horizontally across multiple edge nodes. The mainstream engines currently include Docker Swarm, Mesos, and Kubernetes.

Gao et al. [66] introduced Docker and Docker Swarm into the software definition function of the original Bluetooth, transforming BT-SDF into an extensible and flexible IoT function redefinition framework.

Mesos is a general cluster resource scheduling platform that, together with Marathon, provides all the functionality of a container orchestration engine. Rattihalli et al. [186] presented an Apache-Mesos container migration system based on a two-stage elastic cluster architecture, showing how this framework could be used to manage a scientific workload in three different cloud/cluster environments.

Developed by Google, Kubernetes is an open-source container orchestration engine and supports both Docker and CoreOS containers. Malawski et al. [171] designed an end-to-end containerization solution supporting either Kubernetes or Amazon electrical control system for scientific workflows. The solution can execute mixed workflows, over multiple computing infrastructures if needed, to significantly reduce the burden of infrastructure management.

Kubernetes is associated with some of the more lightweight edge computing platforms, such as KubeEdge and K3s. KubeEdge is an open-source system developed by Huawei for extending container orchestration to edge devices [247]. K3s is a lightweight distributed system launched by Rancher Labs [103]. This product is designed to run Kubernetes in environments with limited resources. K3s supports the x86_64, ARM64, and ARMv7 architectures, allowing K3s to work more flexibly across edge infrastructure. The implementation is a lightweight version of Kubernetes that deletes old, non-essential code, integrates running packaging processes, and uses containers instead of Docker as the runtime container engine. SQLite can also be included as an optional data store.

To solve the problem of data confidentiality, Dupont et al. [55] proposed a migration platform of IoT functions that relies on Kubernetes clusters to perform horizontal roaming and vertical offloading. In a practical demonstration of a healthcare scenario, Dupont and colleagues showed how to use the IoT roaming function with medical data. In a second case, they highlighted the offload capabilities of the platform to optimize the remote diagnosis of mechanical engines.

However, Kaur et al. [120] find that existing Kubernetes solutions lack sufficient robustness to deal with interference and minimize energy consumption in IIoT settings. Therefore, they provided a solution to use an additional competent controller that can manage containers on edge cloud nodes, called KEIDS, which considers a system's carbon footprint and energy consumption and is robust to interference. KEIDS is based on the integer linear programming method of multi-objective optimization problems. Chen et al. [36] proposed a scalable IoT/M2M architecture based on an OpenStack cloud comprising Kubernetes-enabled **Network Function Virtualization (NFV)** plus a management and orchestration engine. Its extended functions are based on Kubernetes with the addition of container-based network functions. Their experiments show that their NFV platform with Kubernetes has better scalability than systems without.

The performance of these different orchestration engines has been compared in multiple studies. Modak et al. [163], for example, compared Docker Swarm and Kubernetes with a cloud-based data security application. Magedanz et al. [84] evaluated how containers on fog nodes influence the performance of various applications, comparing Kubernetes, Mesos-Marathon [160], and Docker Swarm, and how they met the needs of running applications. These researchers also proposed a container orchestration framework for a fog computing infrastructure. Gromann et al. [75] proposed a multi-architecture framework composed of multiple Docker images called PyMon—a Django application that can be used to monitor and collect various data from a host. Before choosing Docker, they compared the resource use of Kubernetes and Docker Swarm.

6 EMERGING FIELDS AND THEIR CHALLENGES IN CDA

While innovation in many ancillary fields will help to accelerate advancements in IIoT, recently emerging technologies in some of these fields show particular promise for helping to promote collaborative data access. Figure 4 provides an overview of these four fields and the potential benefits they may bring. The four fields are 5G, blockchain, semantic web, and machine learning.

6.1 5G

As the communication technology of the next decade, 5G mobile networks provide a significant improvement in speed, latency, capacity, reliability, security, and energy consumption compared to the previous generation of communication technology [38, 111, 184]. 5G defines three key communication scenarios in **enhanced Mobile Broadband (eMBB), massive Machine Type Communication (mMTC)**, and **Ultra-Reliable and Low Latency Communication (URLLC)** [227].

Among them, mMTC has huge benefits for CDA, which can provide access connections around 1k–1M devices/km^2. It is poised to act as an enabler for the connection of sensors, devices, computational platforms, software applications, and operators. Connectivity implies seamless vertical and horizontal integration across all layers of the IIoT architecture. Further, the mobility, flexibility, and cost-effectiveness of 5G can well support CDA, especially in industrial scenarios.

However, the challenges are also in front of CDA to use 5G [4]:

— Gap between 5G and industrial applications. At the current stage, 5G still focuses on eMBB serving for domains like mobile communication, instead of mMTC. Although use cases have been released by 3GPP [2, 3], it is still confusing for manufacturing researchers and engineers to develop 5G applications for CDA.

— Spectrum and operator models. 5G provides the possibility for CDA everywhere through wireless channels, however, to meet the extreme requirements such as low latency and high reliability, a specific spectrum is highly preferred. This can be alternatively replaced by solutions such as regional licenses with well-designed operator models, which is a huge challenge.

— Lack of industrial components and testbeds. Unlike smartphones, industrial devices are significantly different from each other, and integration issues with other existing communication technologies also cannot be ignored. However, the importance of CDA is universality and easy usage, which conflicts with existing issues.

6.2 Machine Learning

There is hope that edge-cloud collaborations will solve many of tomorrow's problems with CAD in the IIoT—especially the resource constraints of the edge and the delays caused by central processing in the cloud [9, 168]. But from real-time task scheduling to computational offloading to security issues to communication delays, the true panacea to all ills are the advances being made every day in machine learning [200].

For task-scheduling problems, machine learning algorithms combined with a good partition strategy can use the operating cycles in edge-cloud architectures to achieve low latency, low energy consumption, and high data throughput [118]. Further, deep learning techniques [143, 151, 213] excel at scheduling predictions, and a well-design neural network can make them quickly [108].

In terms of security issues, machine learning algorithms at the edge can fully protect the access to private data stored on end devices. They can also reduce computing overheads and authentication delays [264].

Communication delays can be overcome with algorithms that reduce the cost of obtaining data from edge devices, decrease bandwidth requirements, and/or reduce waiting times. For example, one of many approaches to delivering these outcomes is to select the most valuable data samples for analysis at the preprocessing stage on the edge node. Cai et al.'s strategy for doing this has an anomaly detection and sample selection model working on each edge node with a machine learning classification scheme that was originally trained on the cloud [31].

Computation offloading is usually designed as a task delegation strategy that distributes jobs to edge devices. Each device is regarded as an agent, and the overall system costs are minimized through machine learning algorithms [110, 138, 144]. Accordingly, edge platforms have blurred into a part of the distributed computing environment to some extent. What sets the two environments apart is the heterogeneity of edge devices and their limited computing and storage resources, which poses a huge challenge to machine learning at the edge [194]. However, edge-cloud architecture, which combines real-time computing on edge devices and global aggregated computing on the cloud, promises a balance between computing power and real-time [49]. Since a typical edge device does not have enough processing power to train a machine learning model in real-time, the model can be generated in the cloud for use at the edge [48, 134, 164, 165, 172].

Machine learning has shown its bright future but still meets the following issues:

— Black box. Neural networks as the main techniques of machine learning are like black boxes, which are almost unexplainable. This issue becomes extremely critical for CDA, since industrial applications prioritize performance factors such as reliability, robustness, and high precision. A black box cannot guarantee these.
— Distributed scenarios. Industrial devices are commonly distributed and collaborated, which are also resource-constrained. This is challenging the mode of edge intelligence. Distributed learning or federated learning [87, 124] is promising, however, they should be more focused on embedded devices for CDA.

6.3 Semantic Web

Semantic web technology is an ever-expanding field that has found its way into more industries over the years. It is a discipline that basically focuses on ways of describing information in a formal and machine-interpretable way [83]. As a facilitator of semantic interoperability, ontologies provide definitions and interpretations of concepts associated with metadata in a particular domain. To name some examples, the **Web Ontology Language (OWL)** is a computational logic-based W3C standard language [156], while the **Semantic Web Rule Language (SWRL)** is designed to express horn-like rules and logic in both the OWL DL and OWL Lite sub-languages of OWL [85]. SPARQL is a resource description framework query language for semantic web databases. These are some of the building blocks at the top levels of semantic technology that are providing users with the necessary structures to link data.

In the near future, IIoT is expected to connect sensors, actuators, devices, gateways, and software services in such massive amounts that the communication loads will be unprecedented and massively challenging. One of the main issues will be handling device heterogeneity [191] and the different standards and formats for storing and communicating data. Semantic technology promises to deal with M2M communications and integrations through its ability to describe objects, share and integrate information, and infer knowledge [72]. Currently, there is a considerable amount of literature on the combination of the semantic web and IoT [131, 218], which is quickly becoming known as the **Semantic Web of Things (SWoT)**. A review of this literature follows.

To facilitate semantic interoperability, IIC publishes the **Industrial Internet Connectivity Framework (IICF)** [112], which redefines the traditional **Open System Interconnection (OSI)** model and implements semantic interoperability on the information layer (the application layer in OSI). In this layer, distributed data management and heterogeneous data interoperability rely on designated ontologies to automatically process and accurately interpret the exchanged data.

Mayer et al. [155] introduced a semantic framework for IIoT called **Open Semantic Framework (OSF)**, in which domain-specific knowledge is grouped in ontologies to enable knowledge integration. The domain-specific ontologies work as a semantic footstone that allows machines to collaborate with each other without human effort. Aimed at the level of interoperability required of Industry 4.0, González et al. proposed the **Standards Ontology (STO)** describing more than 60 I4.0 standards and their relations [74]. This is one of the ontologies that could facilitate the integration of standards on both conceptual and implementational levels. Zhu et al. [270] proposed an ontology-based architecture for integrating semantic information from distributed data nodes into an IIoT system. Similarly, Willner et al. [236] leveraged semantic web technologies to formally describe interactions and exchange information based on the oneM2M standard [217]. Their goal is to ensure that all components in an IIoT system have the ability to exchange information.

Giustozzi et al. [71] developed an ontology of industrial processes to promote knowledge exchange and re-usability from the perspective of context representation and context reasoning to achieve context awareness. Kaed et al. put forward a semantic query engine called SQenIoT [56] and a semantic rule engine named SRE [57] for IIoT gateways, with the aim of retrieving information about the environment to implement dynamic but flexible rule-based control strategies.

Ontology techniques have been an indispensable part of IIoT, while it still meets challenges in practice:

— Domain ontology. Ontology is an important role in every stage of CDA. However, industrial domain knowledge differs much from one to another, so building them is time-consuming. Other techniques such as machine learning should be involved to simplify this work.
— Sharing of ontology. There are some industrial ontologies in the literature, but merging them and sharing the knowledge across the ontology is a challenge. This is important for widely promoting existing ontology to new domains.

6.4 Blockchain

CDA is based on edge-cloud collaboration frameworks, where some thorny issues are rising, including privacy, access threats, and authorization hacks such as modifying key metadata [6, 105]. To ensure the security and privacy of IIoT systems, some are looking to blockchain [199]. Its decentralized nature, resistance to tampering, and strong consistency and traceability make it an efficient and trustworthy method of providing trust between distributed devices [90, 249]. Blockchain consolidates information with a timestamp, then passes itself wholesale to other nodes on the network who verify that the information has been added to the chain through an encrypted hash [18]. This same verifiable technology can be used to build a secure and interoperable ecosystem for edge-cloud collaborative architecture [227]. A universal and decentralized blockchain-enhanced Pub/Sub communication model can circumvent the vulnerabilities publish/subscribe stream models have to malicious attacks and single points of failure [89]. Additionally, the edge-cloud collaborative architecture combined with blockchain technology can be used to share sensitive patient data and perform their calculations [137]. Besides, Zhang et al. proposed the **"Internet of Things Blockchain" (IBoT)** framework [265]. They introduced data encryption in the framework to avoid malicious tampering of credential data and smart contracts to protect data exchange. In contrast, the blockchain-based trusted data management scheme proposed by Ma et al. also incorporates matrix-based multi-channel data segmentation and isolation [266], while Zhang et al. focus more on the importance of the security of shared data in the blockchain industrial IoT domain [263].

The studies on combined blockchain/SDN solutions include Luo et al. [149], who proposed a distributed SDIIoT-enabled blockchain to control different SDNs to synchronize local views between servers and reach a final consensus on the global view. Medhane [157] proposed a security framework that combines blockchain technology with an edge-cloud collaboration architecture and SDNs to face data confidentiality challenges. Guo et al. [77] introduced alliance blockchain and deep reinforcement learning to solve problems of trust and adaptation with resource allocation in an edge-cloud collaboration. Xiao et al. [245] use SDN and blockchain for rapid fake news detection, while Yazdinejad et al. [253] developed a secure and efficient SDN controller architecture supported by blockchain that is provably superior to the classic blockchain.

Blockchain is a hot topic not only for CDA but also for all applications about security, while issues are equally serious in industrial domains:

— High resource consumption. Blockchain takes a lot of computational power, storage, and energy. This is unacceptable for resource-constrained devices in CDA. IoTA is a promising direction to address this issue, which uses a **directed acyclic graph (DAG)** structure. Each

IoTA node does not need to store the entire transaction chain. Moreover, creating and validating transactions on the network does not necessitate validation with every other transaction, leading to a significant reduction in computing and storage requirements. However, industrial applications need long-term tests.

— Delay. The feature of the blockchain is distribution, meanwhile, it produces delay, and this is insufficient for real-time scenarios of CDA, such as an access request of a fast-moving robot.

7 CONCLUSION

Over the past few decades, IIoT has proven to be a highly successful technology in a range of fields. However, as the datasphere approaches the zettabyte scale, the M2M connectivity between the soon-to-be 15 billion heterogeneous devices is challenging conventional data communications and processing systems. This survey focuses on data access as the base of IIoT systems, providing a review of both the existing and emerging technologies enabling collaborative data access. We reviewed the current research on data access from the perspective of standardization, data acquisition, and fusion. To support these aspects, we investigated the history of computing architectures and discussed the techniques for scalable architectures with a focus on edge-cloud collaborations as the trend for future IIoT systems. We also discussed how emerging technologies such as 5G, machine learning, blockchain, and semantic web will advance data access going forward. In summary, the development of IIoT has begun to take shape with numerous achievements in both academic and commercial fields. However, IIoT still faces the challenge of device access—the existing IIoT platforms cannot effectively address the issue of accessing massive and heterogeneous data sources at the edge side. Further research is necessary.

REFERENCES

[1] ISO T. C. 184. 2022. Automation systems and integration. Retrieved from https://www.iso.org/committee/54110.html

[2] 3GPP TR 22.804. 2020. Study on Communication for Automation in Vertical domains. Retrieved from https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3187

[3] 3GPP TR 22.821. 2018. Feasibility Study on LAN Support in 5G. Retrieved from https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3281

[4] 5G ACIA. 2019. 5G for Connected Industries and Automation. Retrieved from https://5g-acia.org/wp-content/uploads/2021/04/WP_5G_for_Connected_Industries_and_Automation_Download_19.03.19.pdf

[5] ABB. 2022. Introducing ABB Ability. Retrieved from https://global.abb/topic/ability/en/about

[6] Giuseppe Aceto, Valerio Persico, and Antonio Pescapé. 2019. A survey on information and communication technologies for industry 4.0: State-of-the-art, taxonomies, perspectives, and challenges. *IEEE Commun. Surv. Tutor.* 21, 4 (2019), 3467–3501.

[7] Tolulope Adesina and Oladiipo Osasona. 2019. A novel cognitive IoT gateway framework: Towards a holistic approach to IoT interoperability. In *IEEE 5th World Forum on Internet of Things (WF-IoT'19)*. IEEE, 53–58.

[8] A. Ahmed and G. Pierre. 2018. Docker container deployment in fog computing infrastructures. In *IEEE International Conference on Edge Computing (EDGE'18)*. 1–8.

[9] Khaled Al-Gumaei, Kornelia Schuba, Andrej Friesen, Sascha Heymann, Carsten Pieper, Florian Pethig, and Sebastian Schriegel. 2018. A survey of internet of things and big data integrated solutions for industrie 4.0. In *A Survey of Internet of Things and Big Data Integrated Solutions for Industrie 4.0*, Vol. 1. IEEE, 1417–1424.

[10] Fadi Al-Turjman and Sinem Alturjman. 2020. 5G/IoT-enabled UAVs for multimedia delivery in industry-oriented applications. *Multim. Tools Applic.* 79 (2020), 8627–8648.

[11] Damminda Alahakoon and Xinghuo Yu. 2015. Smart electricity meter data intelligence for future energy systems: A survey. *IEEE Trans. Industr. Inform.* 12, 1 (2015), 425–436.

[12] Mary B. Alatise and Gerhard P. Hancke. 2017. Pose estimation of a mobile robot based on fusion of IMU data and vision data using an extended Kalman filter. *Sensors* 17, 10 (2017), 2164.

[13] Amazon. 2023. AWS IoT overview. Retrieved from https://aws.amazon.com/iot/?nc1=h_ls

[14] Babu S. Anish, Hareesh M. J., John Paul Martin, Sijo Cherian, and Yedhu Sastri. 2014. System performance evaluation of para virtualization, container virtualization, and full virtualization using Xen, OpenVZ, and XenServer. In *4th International Conference on Advances in Computing and Communications*. 247–250.

[15] Ansif Arooj, Muhammad Shoaib Farooq, Tariq Umer, Ghulam Rasool, and Bo Wang. 2020. Cyber physical and social networks in IoV (CPSN-IoV): A multimodal architecture in edge-based networks for optimal route selection using 5G technologies. *IEEE Access* 8 (2020), 33609–33630.

[16] Md. Asif-Ur-Rahman, Fariha Afsana, Mufti Mahmud, M. Shamim Kaiser, Muhammad R. Ahmed, Omprakash Kaiwartya, and Anne James-Taylor. 2018. Toward a heterogeneous mist, fog, and cloud-based framework for the internet of healthcare things. *IEEE Internet Things J.* 6, 3 (2018), 4049–4062.

[17] Hany F. Atlam, Robert J. Walters, and Gary B. Wills. 2018. Fog computing and the internet of things: A review. *Big Data Cogn. Comput.* 2, 2 (2018), 10.

[18] Arshdeep Bahga and Vijay K. Madisetti. 2016. Blockchain platform for industrial internet of things. *J. Softw. Eng. Applic.* 9, 10 (2016), 533–546.

[19] Roger Baig, Roger Pueyo Centelles, Felix Freitag, and Leandro Navarro. 2017. On edge microclouds to provide local container-based services. In *Global Information Infrastructure and Networking Symposium (GIIS'17)*. 31–36.

[20] Balena. Balena. Retrieved from https://www.balena.io/

[21] Suprateek Banerjee and Daniel Großmann. 2016. An electronic device description language based approach for communication with DBMS and file system in an industrial automation scenario. In *IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA'16)*. IEEE, 1–4.

[22] Shuang Bao, Hairong Yan, Qingping Chi, Zhibo Pang, and Yuying Sun. 2017. FPGA-based reconfigurable data acquisition system for industrial sensors. *IEEE Trans. Industr. Inform.* 13, 4 (2017), 1503–1512.

[23] A. Anand Bardwaj, M Anandaraj, K. Kapil, S. Vasuhi, and V. Vaidehi. 2008. Multi sensor data fusion methods using sensor data compression and estimated weights. In *International Conference on Signal Processing, Communications and Networking*. IEEE, 250–254.

[24] Luis Barreto, Antonio Amaral, and Teresa Pereira. 2017. Industry 4.0 implications in logistics: An overview. *Procedia Manuf.* 13 (2017), 1245–1252.

[25] Flavio Bonomi, Rodolfo Milito, Jiang Zhu, and Sateesh Addepalli. 2012. Fog computing and its role in the internet of things. In *1st Edition of the MCC Workshop on Mobile Cloud Computing*. 13–16.

[26] Borja Bordel, Diego Sánchez De Rivera, and Ramón Alcarria. 2016. Plug-and-play transducers in cyber-physical systems for device-driven applications. In *10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS'16)*. IEEE, 316–321.

[27] Mike Botts, George Percivall, Carl Reed, and John Davidson. 2006. OGC® sensor web enablement: Overview and high level architecture. In *2nd International Conference on GeoSensor Networks*. Springer, 175–190.

[28] Hugh Boyes, Bil Hallaq, Joe Cunningham, and Tim Watson. 2018. The Industrial Internet of Things (IIoT): An analysis framework. *Comput. Industr.* 101 (2018), 1–12.

[29] Arne Bröring, Patrick Maué, Krzysztof Janowicz, Daniel Nüst, and Christian Malewski. 2011. Semantically-enabled sensor plug & play for the sensor web. *Sensors* 11, 8 (2011), 7568–7605.

[30] Edouard Bugnion, Scott Devine, Mendel Rosenblum, Jeremy Sugerman, and Edward Y. Wang. 2012. Bringing virtualization to the X86 architecture with the original VMware workstation. *ACM Trans. Comput. Syst.* 30, 4 (2012), 1–51.

[31] He Cai, Cunqing Hua, and Wenchao Xu. 2019. Design of active learning framework for collaborative anomaly detection. In *11th International Conference on Wireless Communications and Signal Processing (WCSP'19)*. IEEE, 1–7.

[32] Ronghui Cao, Zhuo Tang, Chubo Liu, and Bharadwaj Veeravalli. 2020. A scalable multicloud storage architecture for cloud-supported medical internet of things. *IEEE Internet Things J.* 7, 3 (2020), 1641–1654.

[33] Hyunseok Chang, Adiseshu Hari, Sarit Mukherjee, and T. V. Lakshman. 2014. Bringing the cloud to the edge. In *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS'14)*. IEEE, 346–351.

[34] Wo L. Chang, David Boyd, and Orit Levin. 2019. NIST Big Data Interoperability Framework: Volume 6, Reference Architecture. (October 2019).

[35] Djabir Abdeldjalil Chekired, Lyes Khoukhi, and Hussein T. Mouftah. 2018. Industrial IoT data scheduling based on hierarchical fog computing: A key for enabling smart factory. *IEEE Trans. Industr. Inform.* 14, 10 (2018), 4590–4602.

[36] Hung-Li Chen and Fuchun Joseph Lin. 2019. Scalable IoT/M2M platforms based on Kubernetes-enabled NFV MANO architecture. In *International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. 1106–1111.

[37] Jiasi Chen and Xukan Ran. 2019. Deep learning with edge computing: A review. *Proc. IEEE* 107, 8 (2019), 1655–1674.

[38] Jiangfeng Cheng, Weihai Chen, Fei Tao, and Chun-Liang Lin. 2018. Industrial IoT in 5G environment towards smart manufacturing. *J. Industr. Inf. Integ.* 10 (2018), 10–19.

[39] Ledan Cheng, Songtao Guo, Ying Wang, and Yuanyuan Yang. 2016. Lifting wavelet compression based data aggregation in big data wireless sensor networks. In *IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS'16)*. IEEE, 561–568.

[40] Ying Cheng, Yongping Zhang, Ping Ji, Wenjun Xu, Zude Zhou, and Fei Tao. 2018. Cyber-physical integration for moving digital factories forward towards smart manufacturing: A survey. *Int. J. Advan. Manuf. Technol.* 97, 1–4 (2018), 1209–1221.

[41] Avarachan Cherian, Darold Wobschall, and Mehrdad Sheikholeslami. 2017. An IoT interface for industrial analog sensor with IEEE 21451 protocol. In *IEEE Sensors Applications Symposium (SAS'17)*. IEEE, 1–5.

[42] Qingping Chi, Hairong Yan, Chuan Zhang, Zhibo Pang, and Li Da Xu. 2014. A reconfigurable smart sensor interface for industrial WSN in IoT environment. *IEEE Trans. Industr. Inform.* 10, 2 (2014), 1417–1425.

[43] Yuanfang Chi, Yanjie Dong, Jane Wang, F. Richard Yu, and Victor C. M. Leung. 2022. Knowledge-based fault diagnosis in industrial internet of things: A survey. *IEEE Internet Things J.* 9, 15 (2022), 12886–12900.

[44] Industrial Internet Consortium. 2022. Industrial internet reference architecture. Retrieved from https://www.iiconsortium.org/IIRA/

[45] Li Da Xu, Wu He, and Shancang Li. 2014. Internet of things in industries: A survey. *IEEE Trans. Industr. Inform.* 10, 4 (2014), 2233–2243.

[46] Jad Darrous, Thomas Lambert, and Shadi Ibrahim. 2019. On the importance of container image placement for service provisioning in the edge. In *28th International Conference on Computer Communication and Networks (ICCCN'19)*. 1–9.

[47] Suparna De, Maria Bermudez-Edo, Honghui Xu, and Zhipeng Cai. 2022. Deep generative models in the industrial internet of things: A survey. *IEEE Trans. Industr. Inform.* 18, 9 (2022), 5728–5737.

[48] Shuiguang Deng, Zhengzhe Xiang, Peng Zhao, Javid Taheri, Honghao Gao, Jianwei Yin, and Albert Y. Zomaya. 2020. Dynamical resource allocation in edge for trustable internet-of-things systems: A reinforcement learning method. *IEEE Trans. Industr. Inform.* 16, 9 (2020), 6103–6113.

[49] Shuiguang Deng, Hailiang Zhao, Weijia Fang, Jianwei Yin, Schahram Dustdar, and Albert Y. Zomaya. 2020. Edge intelligence: The confluence of edge computing and artificial intelligence. *IEEE Internet Things J.* 7, 8 (2020), 7457–7469.

[50] Christian Diedrich, Alexander Belyaev, Tizian Schröder, Jens Vialkowitsch, Alexander Willmann, Thomas Usländer, Heiko Koziolek, Jörg Wende, Florian Pethig, and Oliver Niggemann. 2017. Semantic interoperability for asset communication within smart factories. In *22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA'17)*. IEEE, 1–8.

[51] Zhiming Ding, Bin Yang, Yuanying Chi, and Limin Guo. 2015. Enabling smart transportation systems: A parallel spatio-temporal database approach. *IEEE Trans. Comput.* 65, 5 (2015), 1377–1391.

[52] Jasenka Dizdarević, Francisco Carpio, Admela Jukan, and Xavi Masip-Bruin. 2019. A survey of communication protocols for internet of things and related challenges of fog and cloud computing integration. *ACM Computing Surveys (CSUR)* 51, 6 (2019), 1–29.

[53] Docker. What is a Container? Retrieved from https://www.docker.com/resources/what-container. (n.d.).

[54] Rong Du, Paolo Santi, Ming Xiao, Athanasios V. Vasilakos, and Carlo Fischione. 2018. The sensable city: A survey on the deployment and management for smart city monitoring. *IEEE Commun. Surv. Tutor.* 21, 2 (2018), 1533–1560.

[55] Corentin Dupont, Raffaele Giaffreda, and Luca Capra. 2017. Edge computing in IoT context: Horizontal and vertical Linux container migration. In *2017 Global Internet of Things Summit (GIoTS)*. IEEE, 1–4.

[56] Charbel El Kaed, Imran Khan, Hicham Hossayni, and Philippe Nappey. 2016. SQenloT: Semantic query engine for industrial Internet-of-Things gateways. In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*. IEEE, 204–209.

[57] Charbel El Kaed, Imran Khan, Andre Van Den Berg, Hicham Hossayni, and Christophe Saint-Marcel. 2017. SRE: Semantic rules engine for the industrial Internet-of-Things gateways. *IEEE Trans. Industr. Inform.* 14, 2 (2017), 715–724.

[58] Melike Erol-Kantarci and Hussein T. Mouftah. 2014. Energy-efficient information and communication infrastructures in the smart grid: A survey on interactions and open issues. *IEEE Commun. Surv. Tutor.* 17, 1 (2014), 179–197.

[59] Alireza Esfahani, Georgios Mantas, Rainer Matischek, Firooz B Saghezchi, Jonathan Rodriguez, Ani Bicaku, Silia Maksuti, Markus G. Tauber, Christoph Schmittner, and Joaquim Bastos. 2017. A lightweight authentication mechanism for M2M communications in industrial IoT environment. *IEEE Internet Things J.* 6, 1 (2017), 288–296.

[60] Lennart Espe, Anshul Jindal, Vladimir Podolskiy, and Michael Gerndt. 2020. Performance Evaluation of Container Runtimes. In *Proceedings of the 10th International Conference on Cloud Computing and Services Science (CLOSER'20)*. 273–281.

[61] Xi Fang, Satyajayant Misra, Guoliang Xue, and Dejun Yang. 2011. Smart grid—The new and improved power grid: A survey. *IEEE Commun. Surv. Tutor.* 14, 4 (2011), 944–980.

[62] Miguel A. Fernandes, Samuel G. Matos, Emanuel Peres, Carlos R. Cunha, Juan A. López, P. J. S. G. Ferreira, M. J. C. S. Reis, and Raul Morais. 2013. A framework for wireless sensor networks management for precision viticulture and agriculture based on IEEE 1451 standard. *Comput. Electron. Agric.* 95 (2013), 19–30.

[63] Santiago Figueroa-Lorenzo, Javier Añorga, and Saioa Arrizabalaga. 2020. A survey of IIoT protocols: A measure of vulnerability risk analysis based on CVSS. *ACM Comput. Surv.* 53, 2 (2020), 1–53.

[64] Ian Foster. 2003. THE GRID: Computing without bounds. *Scient. Am.* 288, 4 (2003), 78–85.

[65] Paula Fraga-Lamas, Tiago M. Fernández-Caramés, and Luis Castedo. 2017. Towards the internet of smart trains: A review on industrial IoT-connected railways. *Sensors* 17, 6 (2017), 1457.

[66] Yuan Gao, Haoxuan Wang, and Xin Huang. 2016. Applying Docker swarm cluster into software defined internet of things. In *8th International Conference on Information Technology in Medicine and Education (ITME'16).* IEEE, 445–449.

[67] Rafael Garcia, Ann Gordon-Ross, and Alan D. George. 2009. exploiting partially reconfigurable FPGAs for situation-based reconfiguration in wireless sensor networks. In *17th IEEE Symposium on Field Programmable Custom Computing Machines.* IEEE, 243–246.

[68] Dimitrios Georgakopoulos, Prem Prakash Jayaraman, Maria Fazia, Massimo Villari, and Rajiv Ranjan. 2016. Internet of things and edge cloud computing roadmap for manufacturing. *IEEE Cloud Comput.* 3, 4 (2016), 66–73.

[69] Ammar Gharaibeh, Mohammad A. Salahuddin, Sayed Jahed Hussini, Abdallah Khreishah, Issa Khalil, Mohsen Guizani, and Ala Al-Fuqaha. 2017. Smart cities: A survey on data management, security, and enabling technologies. *IEEE Commun. Surv. Tutor.* 19, 4 (2017), 2456–2501.

[70] Nam Ky Giang, Victor C. M. Leung, and Rodger Lea. 2016. On developing smart transportation applications in fog computing paradigm. In *6th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications.* 91–98.

[71] Franco Giustozzi, Julien Saunier, and Cecilia Zanni-Merk. 2018. Context modeling for Industry 4.0: An ontology-based proposal. *Procedia Comput. Sci.* 126 (2018), 675–684.

[72] Sotirios K. Goudos, Panagiotis I. Dallas, Stella Chatziefthymiou, and Sofoklis Kyriazakos. 2017. A survey of IoT key enabling and future technologies: 5G, mobile IoT, sematic web and applications. *Wirel. Person. Commun.* 97, 2 (2017), 1645–1675.

[73] Robert David Graham and Peter C. Johnson. 2014. Finite state machine parsing for internet protocols: Faster than you think. In *IEEE Security and Privacy Workshops.* IEEE, 185–190.

[74] Irlán Grangel-González, Paul Baptista, Lavdim Halilaj, Steffen Lohmann, Maria-Esther Vidal, Christian Mader, and Sören Auer. 2017. The Industry 4.0 standards landscape from a semantic integration perspective. In *22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA'17).* IEEE, 1–8.

[75] Marcel Großmann and Clemens Klug. 2017. Monitoring container services at the network edge. In *29th International Teletraffic Congress (ITC'17)..* 130–133.

[76] Jean A. Guevara, Enrique A. Vargas, Arturo F. Fatecha, and Federico Barrero. 2015. Dynamically reconfigurable WSN node based on ISO/IEC/IEEE 21451 TEDS. *IEEE Sensors J.* 15, 5 (2015), 2567–2576.

[77] Shaoyong Guo, Yao Dai, Siya Xu, Xuesong Qiu, and Feng Qi. 2019. Trusted cloud-edge network resource management: Drl-driven service function chain orchestration for IoT. *IEEE Internet of Things Journal* 7, 7 (2019).

[78] Quang Phuc Ha, Santanu Metia, and Manh Duong Phung. 2020. Sensing data fusion for enhanced indoor air quality monitoring. *IEEE Sensors J.* 20, 8 (2020), 4430–4441.

[79] Lida Haghnegahdar, Sameehan S. Joshi, and Narendra B. Dahotre. 2022. From IoT-based cloud manufacturing approach to intelligent additive manufacturing: Industrial Internet of Things—An overview. *Int. J. Advan. Manuf. Technol.* 79 (2022), 1–18.

[80] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan. 2015. The rise of "big data" on cloud computing: Review and open research issues. *Inf. Syst.* 47 (2015), 98–115.

[81] Abhishek Hazra, Mainak Adhikari, Tarachand Amgoth, and Satish Narayana Srirama. 2021. A comprehensive survey on interoperability for IIoT: Taxonomy, standards, and future directions. *ACM Comput. Surv.* 55, 1 (2021), 1–35.

[82] Heiko Hinkelmann, Andreas Reinhardt, Sameer Varyani, and Manfred Glesner. 2008. A reconfigurable prototyping platform for smart sensor networks. In *4th Southern Conference on Programmable Logic.* 125–130.

[83] Pascal Hitzler, Markus Krotzsch, and Sebastian Rudolph. 2009. *Foundations of Semantic Web Technologies.* CRC Press.

[84] Saiful Hoque, Mathias Santos De Brito, Alexander Willner, Oliver Keil, and Thomas Magedanz. 2017. Towards container orchestration in fog computing infrastructures. In *IEEE 41st Annual Computer Software and Applications Conference (COMPSAC'17).* 294–299.

[85] Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, and Mike Dean. 2004. SWRL: A semantic web rule language combining OWL and RuleML. *W3C Memb. Submiss.* 21, 79 (2004), 1–31.

[86] M. Shamim Hossain and Ghulam Muhammad. 2016. Cloud-assisted Industrial Internet of Things (IIoT)–enabled framework for health monitoring. *Comput. Netw.* 101 (2016), 192–202.

[87] Kevin Hsieh, Aaron Harlap, Nandita Vijaykumar, Dimitris Konomis, Gregory R. Ganger, Phillip B. Gibbons, and Onur Mutlu. 2017. Gaia: Geo-distributed machine learning approaching LAN speeds. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI'17).* 629–647.

[88] Yu-Chen Hsieh, Hua-Jun Hong, Pei-Hsuan Tsai, Yu-Rong Wang, Qiuxi Zhu, M.d. Yusuf Sarwar Uddin, Nalini Venkatasubramanian, and Cheng-Hsin Hsu. 2018. Managed edge computing on Internet-of-Things devices for smart city applications. In *IEEE/IFIP Network Operations and Management Symposium.* 1–2.

[89] Bobo Huang, Rui Zhang, Zhihui Lu, Yiming Zhang, Jie Wu, Lu Zhan, and Patrick C. K. Hung. 2020. BPS: A reliable and efficient pub/sub communication model with blockchain-enhanced paradigm in multi-tenant edge cloud. *Journal of Parallel and Distributed Computing* 143, (2020).

[90] Junqin Huang, Linghe Kong, Guihai Chen, Min-You Wu, Xue Liu, and Peng Zeng. 2019. Towards secure industrial IoT: Blockchain system with credit-based consensus mechanism. *IEEE Trans. Industr. Inform.* 15, 6 (2019), 3680–3689.

[91] Xumin Huang, Peichun Li, and Rong Yu. 2019. Social welfare maximization in container-based task scheduling for parked vehicle edge computing. *IEEE Commun. Lett.* 23, 8 (2019), 1347–1351.

[92] Yuzhou Huang, Kaiyu Cai, Ran Zong, and Yugang Mao. 2019. Design and implementation of an edge computing platform architecture using Docker and Kubernetes for machine learning. In *3rd International Conference on High Performance Compilation, Computing and Communications.* 29–32.

[93] Ru Huo, Shiqin Zeng, Zhihao Wang, Jiajia Shang, Wei Chen, Tao Huang, Shuo Wang, F. Richard Yu, and Yunjie Liu. 2022. A comprehensive survey on blockchain in industrial internet of things: Motivations, research progresses, and future challenges. *IEEE Commun. Surv. Tutor.* 24, 1 (2022), 88–122.

[94] Joo-Young Hwang, Sang-Bum Suh, Sung-Kwan Heo, Chan-Ju Park, Jae-Min Ryu, Seong-Yeol Park, and Chul-Ryun Kim. 2008. Xen on ARM: System virtualization using Xen hypervisor for ARM-based secure mobile phones. In *5th IEEE Consumer Communications and Networking Conference.* 257–261.

[95] IEC-International Electrotechnical Commission and others. 2016. IEC 61360-6. Retrieved from https://webstore.iec.ch/publication/25984

[96] ISO. ISO/IEC 20005:2013. Retrieved from https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/09/50952.html

[97] ISO. ISO/IEC 20248:2018. Retrieved from https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/74/67412.html

[98] ISO. ISO/IEC 29143:2011. Retrieved from https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/04/51/45166.html

[99] ISO. ISO/IEC 29175:2012. Retrieved from https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/04/52/45254.html

[100] ISO. ISO/IEC 29182-1:2013. Retrieved from https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/04/52/45261.html

[101] ISO. ISO/IEC 30101:2014. Retrieved from https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/32/53221.html

[102] Davood Izadi, Jemal H. Abawajy, Sara Ghanavati, and Tutut Herawan. 2015. A data fusion method in wireless sensor networks. *Sensors* 15, 2 (2015), 2964–2979.

[103] Kavita Jaiswal, Srichandan Sobhanayak, Bhabendu Kumar Mohanta, and Debasish Jena. 2017. IoT-cloud based framework for patient's data collection in smart healthcare system using Raspberry-Pi. In *International Conference on Electrical and Computing Technologies and Applications (ICECTA'17).* IEEE, 1–4.

[104] Bilal Jan, Haleem Farman, Murad Khan, Muhammad Talha, and Ikram Ud Din. 2019. Designing a smart transportation system: An internet of things and big data approach. *IEEE Wirel. Commun.* 26, 4 (2019), 73–79.

[105] Kanwal Janjua, Munam Ali Shah, Ahmad Almogren, Hasan Ali Khattak, Carsten Maple, and Ikram Ud Din. 2020. Proactive forensics in IoT: Privacy-aware log-preservation architecture in fog-enabled-cloud using holochain and containerization technologies. *Electronics* 9, 7 (2020), 1172.

[106] Juergen Jasperneite, Thilo Sauter, and Martin Wollschlaeger. 2020. Why we need automation models: Handling complexity in Industry 4.0 and the internet of things. *IEEE Industr. Electron. Mag.* 14, 1 (Mar. 2020), 29–40.

[107] Isam Mashhour Al Jawarneh, Paolo Bellavista, Filippo Bosi, Luca Foschini, Giuseppe Martuscelli, Rebecca Montanari, and Amedeo Palopoli. 2019. Container orchestration engines: A thorough functional and performance comparison. In *IEEE International Conference on Communications (ICC'19).* 1–6.

[108] Chengfeng Jian, Jing Ping, and Meiyu Zhang. 2021. A cloud edge-based two-level hybrid scheduling learning model in cloud manufacturing. *Int. J. Product. Res.* 59, 16 (2021), 4836–4850.

[109] Bin Jiang, Jianqiang Li, Guanghui Yue, and Houbing Song. 2021. Differential privacy for industrial internet of things: Opportunities, applications, and challenges. *IEEE Internet Things J.* 8, 13 (2021), 10430–10451.

[110] Feibo Jiang, Kezhi Wang, Li Dong, Cunhua Pan, Wei Xu, and Kun Yang. 2020. Deep-learning-based joint resource scheduling algorithms for hybrid MEC networks. *IEEE Internet Things J.* 7, 7 (2020), 6252–6265.

[111] Tao Jiang, Jianhua Zhang, Pan Tang, Lei Tian, Yi Zheng, Jianwu Dou, Henrik Asplund, Leszek Raschkowski, Raffaele D'Errico, and Tommi Jämsä. 2021. 3GPP standardized 5G channel model for IIoT scenarios: A survey. *IEEE Internet Things J.* 8, 11 (2021), 8799–8815.

[112] Rajive Joshi, Paul Didier, Jaime Jimenez, and Timothy Carey. 2017. The industrial internet of things volume G5: Connectivity framework. *Industr. Internet Consort. Rep.* (2017). https://www.iiconsortium.org/pdf/IIC_PUB_G5_V1.0_PB_20170228.pdf

[113] Amin Jula, Elankovan Sundararajan, and Zalinda Othman. 2014. Cloud computing service composition: A systematic literature review. *Expert Syst. Applic.* 41, 8 (2014), 3809–3824.

[114] Jun Zhang, Kai Chen, Baojing Zuo, Ruhui Ma, Yaozu Dong, and Haibing Guan. 2010. Performance analysis towards a KVM-based embedded real-time virtualization architecture. In *5th International Conference on Computer Sciences and Convergence Information Technology*. 421–426.

[115] Shahidullah Kaiser, M.d. Sadun Haq, Ali Şaman Tosun, and Turgay Korkmaz. 2022. Container technologies for ARM architecture: A comprehensive survey of the state-of-the-art. *IEEE Access* 10 (2022), 84853–84881.

[116] Syed Rameez Ullah Kakakhel, Lauri Mukkala, Tomi Westerlund, and Juha Plosila. 2018. Virtualization at the network edge: A technology perspective. In *3rd International Conference on Fog and Mobile Edge Computing (FMEC'18)*. 87–92.

[117] David Kampert and Ulrich Epple. 2012. Modeling asset information for interoperable software systems. In *IEEE 10th International Conference on Industrial Informatics*. IEEE, 947–952.

[118] Yiping Kang, Johann Hauswald, Cao Gao, Austin Rovinski, Trevor Mudge, Jason Mars, and Lingjia Tang. 2017. Neurosurgeon: Collaborative intelligence between the cloud and mobile edge. *ACM SIGARCH Comput. Archit. News* 45, 1 (2017), 615–629.

[119] Gayatri Kapil, Alka Agrawal, and R. A. Khan. 2016. A study of big data characteristics. In *International Conference on Communication and Electronics Systems (ICCES'16)*. 1–4.

[120] Kuljeet Kaur, Sahil Garg, Georges Kaddoum, Syed Hassan Ahmed, and Mohammed Atiquzzaman. 2020. KEIDS: Kubernetes based Energy and Interference Driven Scheduler for industrial IoT in edge-cloud ecosystem. *IEEE Internet Things J.* 7, 5 (2020), 4228–4237.

[121] Ruhul Amin Khalil, Nasir Saeed, Mudassir Masood, Yasaman Moradi Fard, Mohamed-Slim Alouini, and Tareq Y. Al-Naffouri. 2021. Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications. *IEEE Internet Things J.* 8, 14 (2021), 11016–11040.

[122] Wazir Zada Khan, M. H. Rehman, Hussein Mohammed Zangoti, Muhammad Khalil Afzal, Nasrullah Armi, and Khaled Salah. 2020. Industrial internet of things: Recent advances, enabling technologies and open challenges. *Comput. Electric. Eng.* 81 (2020), 106522.

[123] Jayoung Kim, Alan S. Campbell, Berta Esteban-Fernández de Ávila, and Joseph Wang. 2019. Wearable biosensors for healthcare monitoring. *Nat. Biotechnol.* 37, 4 (2019), 389–406.

[124] Tim Kraska, Ameet Talwalkar, John C. Duchi, Rean Griffith, Michael J. Franklin, and Michael I. Jordan. 2013. MLbase: A Distributed Machine-learning System. In *6th Biennial Conference on Innovative Data Systems Research (CIDR'13)*. 2–1.

[125] X. Krasniqi and E. Hajrizi. 2016. Use of IoT technology to drive the automotive industry from connected to full autonomous vehicles. *IFAC-PapersOnLine* 49, 29 (2016), 269–274.

[126] Yana Esteves Krasteva, Jorge Portilla, Eduardo de la Torre, and Teresa Riesgo. 2011. Embedded runtime reconfigurable nodes for wireless sensor networks applications. *IEEE Sensors J.* 11, 9 (2011), 1800–1810.

[127] Nihal Kularatna and B. H. Sudantha. 2008. An environmental air pollution monitoring system based on the IEEE 1451 standard for low cost requirements. *IEEE Sensors J.* 8, 4 (2008), 415–422.

[128] Anuj Kumar and Gerhard P. Hancke. 2014. An energy-efficient smart comfort sensing system based on the IEEE 1451 standard for green buildings. *IEEE Sensors J.* 14, 12 (2014), 4245–4252.

[129] Anuj Kumar, Abhishek Singh, Ashok Kumar, Manoj Kumar Singh, Pinakeswar Mahanta, and Subhas Chandra Mukhopadhyay. 2018. Sensing technologies for monitoring intelligent buildings: A review. *IEEE Sensors J.* 18, 12 (2018), 4847–4860.

[130] Anuj Kumar, I. P. Singh, and S. K. Sud. 2011. Energy efficient and low-cost indoor environment monitoring system based on the IEEE 1451 standard. *IEEE Sensors J.* 11, 10 (2011), 2598–2610.

[131] K. N. Prashanth Kumar, V. Ravi Kumar, and K. Raghuveer. 2017. A survey on semantic web technologies for the internet of things. In *International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC'17)*. IEEE, 316–322.

[132] Benoit Larras and Antoine Frappé. 2020. Distributed clique-based neural networks for data fusion at the edge. In *IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS'20)*. 55–58.

[133] Billy Pik Lik Lau, Sumudu Hasala Marakkalage, Yuren Zhou, Naveed Ul Hassan, Chau Yuen, Meng Zhang, and U-Xuan Tan. 2019. A survey of data fusion in smart city applications. *Inf. Fusion* 52 (2019), 357–374.

[134] Juyong Lee and Jihoon Lee. 2020. Juice recipe recommendation system using machine learning in MEC environment. *IEEE Consum. Electron. Mag.* 9, 5 (2020), 79–84.

[135] Kyung Chang Lee, Man Ho Kim, Suk Lee, and Hong Hee Lee. 2004. IEEE-1451-based smart module for in-vehicle networking systems of intelligent vehicles. *IEEE Trans. Industr. Electron.* 51, 6 (2004), 1150–1158.

[136] Joerg Leukel. 2004. Standardization of product ontologies in B2B relationships-on the role of ISO 13584. In *Americas Conference on Information Systems (AMCIS'04)*. 510. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=2086&context=amcis2004

[137] Peilong Li, Chen Xu, Hao Jin, Chunyang Hu, Yan Luo, Yu Cao, Jomol Mathew, and Yunsheng Ma. 2020. ChainSDI: A software-defined infrastructure for regulation-compliant home-based healthcare services secured by blockchains. *IEEE Syst. J.* 14, 2 (2020), 2042–2053.

[138] Quanyi Li, Haipeng Yao, Tianle Mai, Chunxiao Jiang, and Yan Zhang. 2020. Reinforcement-learning- and belief-learning-based double auction mechanism for edge computing resource allocation. *IEEE Internet Things J.* 7, 7 (2020), 5976–5985.

[139] Youhuizi Li, Jiancheng Zhang, Congfeng Jiang, Jian Wan, and Zujie Ren. 2019. PINE: Optimizing performance isolation in container environments. *IEEE Access* 7 (2019), 30410–30422.

[140] Torben R. Licht. 2001. The IEEE 1451.4 proposed standard. *IEEE Instrum. Measur. Mag.* 4, 1 (2001), 12–18.

[141] Alex X. Liu, Chad R. Meiners, Eric Norige, and Eric Torng. 2014. High-speed application protocol parsing and extraction for deep flow inspection. *IEEE J. Select. Areas Commun.* 32, 10 (2014), 1864–1880.

[142] Boyi Liu, Lujia Wang, and Ming Liu. 2019. Lifelong Federated Reinforcement Learning: A learning architecture for navigation in cloud robotic systems. *IEEE Robotics and Automation Letters* 4, 4 (2019), 4555–4562.

[143] Fanzhen Liu, Shan Xue, Jia Wu, Chuan Zhou, Wenbin Hu, Cecile Paris, Surya Nepal, Jian Yang, and Philip S. Yu. 2020. Deep learning for community detection: Progress, challenges and opportunities. In *29th International Joint Conference on Artificial Intelligence (IJCAI'20)*. 4981–4987.

[144] Xiaolan Liu, Jiadong Yu, Jian Wang, and Yue Gao. 2020. Resource allocation with edge computing in IoT networks via machine learning. *IEEE Internet Things J.* 7, 4 (2020), 3415–3426.

[145] Wei Lu, Xianyu Meng, and Guanfei Guo. 2019. Fast service migration method based on virtual machine technology for MEC. *IEEE Internet Things J.* 6, 3 (2019), 4344–4354.

[146] Yang Lu. 2017. Industry 4.0: A survey on technologies, applications and open research issues. *J. Industr. Inf. Integ.* 6 (2017), 1–10.

[147] Yan Lu, Paul Witherell, and Albert Jones. 2020. Standard connections for IIoT empowered smart manufacturing. *Manuf. Lett.* 26 (2020), 17–20.

[148] Henrik Lund, Poul Alberg Østergaard, David Connolly, and Brian Vad Mathiesen. 2017. Smart energy and smart energy systems. *Energy* 137 (2017), 556–565.

[149] Jia Luo, Qianbin Chen, F. Richard Yu, and Lun Tang. 2020. Blockchain-enabled software-defined industrial internet of things with deep reinforcement learning. *IEEE Internet Things J.* 7, 6 (2020), 5466–5480.

[150] Lele Ma, Shanhe Yi, Nancy Carter, and Qun Li. 2019. Efficient live migration of edge services leveraging container layered storage. *IEEE Trans. Mob. Comput.* 18, 9 (2019), 2020–2033.

[151] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Quan Z. Sheng, and Hui Xiong. 2021. A comprehensive survey on graph anomaly detection with deep learning. *arXiv preprint arXiv:2106.07178* (2021).

[152] Mohammed M. Mabkhot, Abdulrahman M. Al-Ahmari, Bashir Salah, and Hisham Alkhalefah. 2018. Requirements of the smart factory system: A survey and perspective. *Machines* 6, 2 (2018), 23.

[153] Andrew Machen, Shiqiang Wang, Kin K. Leung, Bong Jun Ko, and Theodoros Salonidis. 2017. Live service migration in mobile edge clouds. *IEEE Wirel. Commun.* 25, 1 (2017), 140–147.

[154] Ananda Maiti, Alexander A. Kist, and Andrew D. Maxwell. 2018. Automata-based generic model for interoperating context-aware ad-hoc devices in internet of things. *IEEE Internet Things J.* 5, 5 (2018), 3837–3852.

[155] Simon Mayer, Jack Hodges, Dan Yu, Mareike Kritzler, and Florian Michahelles. 2017. An open semantic framework for the industrial internet of things. *IEEE Intell. Syst.* 32, 1 (2017), 96–101.

[156] Deborah L. McGuinness and Frank Van Harmelen. 2004. OWL web ontology language overview. *W3C Recomm.* 10, 10 (2004), 2004.

[157] Darshan Vishwasrao Medhane, Arun Kumar Sangaiah, M. Shamim Hossain, Ghulam Muhammad, and Jin Wang. 2020. Blockchain-enabled distributed security framework for next generation IoT: An edge-cloud and software defined network integrated approach. *IEEE Internet Things J.* 7, 7 (2020), 6143–6149.

[158] Pankaj Mendki. 2018. Docker container based analytics at IoT edge video analytics usecase. In *3rd International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU'18)*. 1–4.

[159] Zhaozong Meng, Zhipeng Wu, Cahyo Muvianto, and John Gray. 2016. A data-oriented M2M messaging mechanism for industrial IoT applications. *IEEE Internet Things J.* 4, 1 (2016), 236–246.

[160] Inc. Mesosphere. 2023. Marathon: A container orchestration platform for Mesos and DC/OS. Retrieved from https://mesosphere.github.io/marathon/

[161] Microsoft. 2023. Azure IoT overview. Retrieved from https://azure.microsoft.com/en-au/overview/iot/#overview

[162] Roberto Minerva, Gyu Myoung Lee, and Noel Crespi. 2020. Digital twin in the IoT context: A survey on technical features, scenarios, and architectural models. *Proc. IEEE* 108, 10 (2020), 1785–1824.

[163] Arsh Modak, S. D. Chaudhary, P. S. Paygude, and S. R. Ldate. 2018. Techniques to secure data on cloud: Docker swarm or Kubernetes? In *2nd International Conference on Inventive Communication and Computational Technologies (ICICCT'18)*. 7–12.

[164] Cristina Morariu, Octavian Morariu, Silviu Răileanu, and Theodor Borangiu. 2020. Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems. *Comput. Industr.* 120 (2020), 103244.
[165] Dariusz Mrozek, Anna Koczur, and Bożena Małysiak-Mrozek. 2020. Fall detection in older adults with mobile IoT devices and machine learning in the cloud and on the edge. *Inf. Sci.* 537 (2020), 132–147.
[166] Luís Neto, Gil Gonçalves, Pedro Torres, Rogério Dionísio, and Sérgio Malhão. 2019. An industry 4.0 self description information model for software components contained in the administration shell. In *8th International Conference on Intelligent Systems and Applications*.
[167] Irene C. L. Ng and Susan Y. L. Wakenshaw. 2017. The internet-of-things: Review and research directions. *Int. J. Res. Market.* 34, 1 (2017), 3–21.
[168] Zhaolong Ning, Xiangjie Kong, Feng Xia, Weigang Hou, and Xiaojie Wang. 2019. Green and sustainable cloud of things: Enabling collaborative edge computing. *IEEE Commun. Mag.* 57, 1 (2019), 72–78.
[169] Masaaki Nishikiori. 2011. Server virtualization with VMware vSphere 4. *Fujitsu Scient. Technic. J.* 47, 3 (2011), 356–361.
[170] OMRON. 2023. Omron CP1W-20EDR1 datasheet. Retrieved from https://datasheet.octopart.com/CP1W-20EDR1-Omron-datasheet-12510914.pdf
[171] Michal Orzechowski, Bartosz Balis, Krystian Pawlik, Maciej Pawlik, and Maciej Malawski. 2018. Transparent deployment of scientific workflows across clouds—Kubernetes approach. In *IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion'18)*. 9–10.
[172] Emmanuel Oyekanlu. 2018. Osmotic collaborative computing for machine learning and cybersecurity applications in industrial IoT networks and cyber physical systems with Gaussian mixture models. In *IEEE 4th International Conference on Collaboration and Internet Computing (CIC'18)*. IEEE, 326–335.
[173] Claus Pahl and Brian Lee. 2015. Containers and clusters for edge cloud architectures—A technology review. In *3rd International Conference on Future Internet of Things and Cloud*. 379–386.
[174] Junmin Park, Hyunjae Park, and Young-June Choi. 2018. Data compression and prediction using machine learning for industrial IoT. In *International Conference on Information Networking (ICOIN'18)*. IEEE, 818–820.
[175] Florian Pethig, Oliver Niggemann, and Armin Walter. 2017. Towards Industrie 4.0 compliant configuration of condition monitoring services. In *IEEE 15th International Conference on Industrial Informatics (INDIN'17)*. IEEE, 271–276.
[176] Alexey S. Petrenko, Sergei A. Petrenko, Krystina A. Makoveichuk, and Petr V. Chetyrbok. 2018. The IIoT/IoT device control model based on narrow-band IoT (NB-IoT). In *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus'18)*. IEEE, 950–953.
[177] Riccardo Petrolo, Valeria Loscri, and Nathalie Mitton. 2017. Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms. *Trans. Emerg. Telecommun. Technol.* 28, 1 (2017), e2931.
[178] Ilia Pietri and Rizos Sakellariou. 2016. Mapping virtual machines onto physical machines in cloud computing: A survey. *Comput. Surv.* 49, 3 (2016).
[179] Jorge Portilla, Teresa Riesgo, and Angel de Castro. 2007. A reconfigurable FPGA-based architecture for modular nodes in wireless sensor networks. In *3rd Southern Conference on Programmable Logic*. 203–206.
[180] D. Potter. 2002. Smart plug and play sensors. *IEEE Instrum. Measur. Mag.* 5, 1 (2002), 28–30.
[181] Gang Qian, Siliang Lu, Donghui Pan, Huasong Tang, Yongbin Liu, and Qunjing Wang. 2019. Edge computing: A promising framework for real-time fault diagnosis and dynamic control of rotating machines using multi-sensor data. *IEEE Sensors J.* 19, 11 (2019), 4211–4220.
[182] Tie Qiu, Ning Chen, Keqiu Li, Mohammed Atiquzzaman, and Wenbing Zhao. 2018. How can heterogeneous internet of things build our future: A survey. *IEEE Commun. Surv. Tutor.* 20, 3 (2018), 2011–2027.
[183] Tie Qiu, Jiancheng Chi, Xiaobo Zhou, Zhaolong Ning, Mohammed Atiquzzaman, and Dapeng Oliver Wu. 2020. Edge computing in industrial internet of things: Architecture, advances and challenges. *IEEE Commun. Surv. Tutor.* 22, 4 (2020), 2462–2488.
[184] Sriganesh K. Rao and Ramjee Prasad. 2018. Impact of 5G technologies on Industry 4.0. *Wirel. Person. Commun.* 100, 1 (2018), 145–159.
[185] M. Mazhar Rathore, Awais Ahmad, Anand Paul, and Gwanggil Jeon. 2015. Efficient graph-oriented smart transportation using internet of things generated big data. In *11th International Conference on Signal-Image Technology & Internet-based Systems (SITIS'15)*. IEEE, 512–519.
[186] Gourav Rattihalli, Madhusudhan Govindaraju, and Devesh Tiwari. 2019. Towards enabling dynamic resource estimation and correction for improving utilization in an Apache Mesos cloud environment. In *19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID'19)*. 188–197.
[187] Partha Pratim Ray. 2016. A survey of IoT cloud platforms. *Fut. Comput. Inform. J.* 1, 1–2 (2016), 35–46.
[188] Yi Ren, Ling Liu, Qi Zhang, Qingbo Wu, Jianbo Guan, Jinzhu Kong, Huadong Dai, and Lisong Shao. 2016. Shared-memory optimizations for inter-virtual-machine communication. *Comput. Surv.* 48, 4 (2016), 1–42.

[189] Yuzheng Ren, Renchao Xie, F. Richard Yu, Tao Huang, and Yunjie Liu. 2020. Potential identity resolution systems for the industrial internet of things: A survey. *IEEE Commun. Surv. Tutor.* 23, 1 (2020), 391–430.

[190] Wenjie Ruan, Quan Z. Sheng, Lina Yao, Xue Li, Nickolas J. G. Falkner, and Lei Yang. 2018. Device-free human localization and tracking with UHF passive RFID tags: A data-driven approach. *J. Netw. Comput. Applic.* 104 (2018), 78–96.

[191] Ahmad-Reza Sadeghi, Christian Wachsmann, and Michael Waidner. 2015. Security and privacy challenges in industrial internet of things. In *52nd Annual Design Automation Conference*. IEEE, 1–6.

[192] T. Sanpechuda and L. Kovavisaruch. 2008. A review of RFID localization: Applications and techniques. In *5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*. IEEE, 769–772.

[193] Mahadev Satyanarayanan. 2017. The emergence of edge computing. *Computer* 50, 1 (2017), 30–39.

[194] Jacqueline Schmitt, Jochen Bönig, Thorbjörn Borggräfe, Gunter Beitinger, and Jochen Deuse. 2020. Predictive model-based quality inspection using machine learning and edge cloud computing. *Adv. Eng. Inform.* 45 (2020), 101101.

[195] Schneider. 2023. TM3XTRA1 remote transmitter module TM3 bus Schneider Electric. Retrieved June 23, 2023 from https://www.schneider-electric.com/en/product/TM3XTRA1/remote-transmitter-module-tm3---bus/

[196] Stan Schneider. 2017. The Industrial Internet of Things (IIoT) applications and taxonomy. *Internet Things Data Analyt. Handb.* (2017), 41–81. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119173601.ch3

[197] Schneider Electric. 2022. What Are EcoStruxure IT Software & Digital Services? Retrieved from https://ecostruxureit.com/what-is-ecostruxure-it/

[198] Karsten Schweichhart. 2016. Reference Architectural Model Industrie 4.0 (RAMI 4.0). Retrieved from https://www.plattform-i40.de

[199] Jayasree Sengupta, Sushmita Ruj, and Sipra Das Bit. 2020. A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT. *J. Netw. Comput. Applic.* 149 (2020), 102481.

[200] Shree Krishna Sharma and Xianbin Wang. 2017. Live data analytics with collaborative edge and cloud processing in wireless IoT networks. *IEEE Access* 5 (2017), 4621–4635.

[201] Kadam Rekha Shashikant and Anju Kulkarni. 2020. Reconfigurable patch antenna design using pin diodes and Raspberry PI for portable device application. *Wirel. Person. Commun.* 112, 3 (2020), 1809–1828.

[202] Mu Shengdong, Xiong Zhengxian, and Tian Yixiang. 2019. Intelligent traffic control system based on cloud computing and big data mining. *IEEE Trans. Industr. Inform.* 15, 12 (2019), 6583–6592.

[203] Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. 2016. Edge computing: Vision and challenges. *IEEE Internet Things J.* 3, 5 (2016), 637–646.

[204] Konstantin Shvachko, Hairong Kuang, Sanjay Radia, and Robert Chansler. 2010. The Hadoop distributed file system. In *IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST'10)*. IEEE, 1–10.

[205] SIEMENS. 2023. Siemens 6ES76470AA001YA2. https://www.dacelsolutions.com/upload/dokumanlar/Siemens%206ES76470AA001YA2.pdf

[206] Siemens. 2022. MindSphere documentation overview. Retrieved from https://siemens.mindsphere.io/en/docs/documentation-overview

[207] Tanwa Sirisakdiwan and Natawut Nupairoj. 2019. Spark framework for real-time analytic of multiple heterogeneous data streams. In *2nd International Conference on Communication Engineering and Technology (ICCET'19)*. IEEE, 1–5.

[208] Emiliano Sisinni, Abusayeed Saifullah, Song Han, Ulf Jennehag, and Mikael Gidlund. 2018. Industrial internet of things: Challenges, opportunities, and directions. *IEEE Trans. Industr. Inform.* 14, 11 (2018), 4724–4734.

[209] Inés Sittón-Candanedo, Ricardo S. Alonso, Juan M. Corchado, Sara Rodríguez-González, and Roberto Casado-Vara. 2019. A review of edge computing reference architectures and a new global edge proposal. *Fut. Gen. Comput. Syst.* 99 (2019), 278–294.

[210] Eugene Y. Song, Martin Burns, Abhinav Pandey, and Thomas Roth. 2019. IEEE 1451 smart sensor digital twin federation for IoT/CPS Research. In *IEEE Sensors Applications Symposium (SAS'19)*. IEEE, 1–6.

[211] Eugene Y. Song and Kang Lee. 2008. Understanding IEEE 1451-networked smart transducer interface standard—What is a smart transducer? *IEEE Instrum. Measur. Mag.* 11, 2 (2008), 11–17.

[212] Dragan H. Stojanović, Natalija M. Stojanović, Igor Đorđević, and Aleksandra I. Stojnev Ilić. 2019. Sensor data fusion and big mobility data analytics for activity recognition. In *14th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS'19)*. IEEE, 66–69.

[213] Xing Su, Shan Xue, Fanzhen Liu, Jia Wu, Jian Yang, Chuan Zhou, Wenbin Hu, Cecile Paris, Surya Nepal, Di Jin, Quan Z. Sheng, and Philip S. Yu. 2022. A comprehensive survey on community detection with deep learning. *IEEE Trans. Neural Netw. Learn. Syst.* (2022), 1–21. https://ieeexplore.ieee.org/document/9732192

[214] Dawei Sun, Vincent C. S. Lee, and Ye Lu. 2016. An intelligent data fusion framework for structural health monitoring. In *IEEE 11th Conference on Industrial Electronics and Applications (ICIEA'16)*. 49–54.

[215] Danfeng Sun, Shan Xue, Huifeng Wu, and Jia Wu. 2021. A data stream cleaning system using edge intelligence for smart city industrial environments. *IEEE Trans. Industr. Inform.* 18, 2 (2021), 1165–1174.
[216] R. Sundaramurthy and V. Nagarajan. 2016. Design and implementation of reconfigurable virtual instruments using Raspberry Pi core. In *International Conference on Communication and Signal Processing (ICCSP'16)*. 2309–2313.
[217] Jorg Swetina, Guang Lu, Philip Jacobs, Francois Ennesser, and JaeSeung Song. 2014. Toward a standardized common M2M service layer platform: Introduction to oneM2M. *IEEE Wirel. Commun.* 21, 3 (2014), 20–26.
[218] Ioan Szilagyi and Patrice Wira. 2016. Ontologies and semantic web for the internet of things—A survey. In *42nd Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 6949–6954.
[219] Seyede Zahra Tajalli, Mohammad Mardaneh, Elaheh Taherian-Fard, Afshin Izadian, Abdollah Kavousi-Fard, Morteza Dabbaghjamanesh, and Taher Niknam. 2020. DoS-resilient distributed optimal scheduling in a fog supporting IIoT-based smart microgrid. *IEEE Trans. Industr. Applic.* 56, 3 (2020), 2968–2977.
[220] Kiyotaka Takahashi, Yuji Ogata, and Youichi Nonaka. 2017. A proposal of unified reference model for smart manufacturing. In *13th IEEE Conference on Automation Science and Engineering (CASE'17)*. 964–969.
[221] Jie Tang, Rao Yu, Shaoshan Liu, and Jean-Luc Gaudiot. 2020. A container based edge offloading framework for autonomous driving. *IEEE Access* 8 (2020), 33713–33726.
[222] Koen Tange, Michele De Donno, Xenofon Fafoutis, and Nicola Dragoni. 2020. A systematic survey of industrial internet of things security: Requirements and fog computing opportunities. *IEEE Commun. Surv. Tutor.* 22, 4 (2020), 2489–2520.
[223] Erdal Tantik and Reiner Anderl. 2017. Integrated data model and structure for the asset administration shell in industrie 4.0. *Procedia CIRP* 60 (2017), 86–91.
[224] Amy J. C. Trappey, Charles V. Trappey, Usharani Hareesh Govindarajan, Allen C. Chuang, and John J. Sun. 2017. A review of essential standards and patent landscapes for the internet of things: A key enabler for industry 4.0. *Adv. Eng. Inform.* 33 (Aug. 2017), 208–229.
[225] D. Ursutiu, C. Samoila, V. Jinga, and F. Altoe. 2016. The future of "hardware–software reconfigurable." In *International Conference on Interactive Collaborative Learning*. Springer, 269–275.
[226] Fredy João Valente, João Paulo Morijo, Kelen Cristiane T. Vivaldini, and Luis Carlos Trevelin. 2019. Fog-based data fusion for heterogeneous IoT sensor networks: A real implementation. In *15th International Conference on Network and Service Management (CNSM'19)*. IEEE, 1–5.
[227] Pal Varga, Jozsef Peto, Attila Franko, David Balla, David Haja, Ferenc Janky, Gabor Soos, Daniel Ficzere, Markosz Maliosz, and Laszlo Toka. 2020. 5G support for industrial IoT applications—Challenges, solutions, and research gaps. *Sensors* 20, 3 (2020), 828.
[228] Jiafu Wan, Shenglong Tang, Di Li, Muhammad Imran, Chunhua Zhang, Chengliang Liu, and Zhibo Pang. 2018. Reconfigurable smart factory for drug packing in healthcare Industry 4.0. *IEEE Trans. Industr. Inform.* 15, 1 (2018), 507–516.
[229] Jiafu Wan, Shenglong Tang, Zhaogang Shu, Di Li, Shiyong Wang, Muhammad Imran, and Athanasios V. Vasilakos. 2016. Software-defined industrial internet of things in the context of industry 4.0. *IEEE Sensors J.* 16, 20 (2016), 7373–7380.
[230] Lan Wang, Shinpei Hayashi, and Motoshi Saeki. 2021. Applying class distance to decide similarity on information models for automated data interoperability. *Int. J. Softw. Eng. Knowl. Eng.* 31, 03 (Mar. 2021), 405–434.
[231] Pan Wang, Feng Ye, and Xuejiao Chen. 2018. A smart home gateway platform for data collection and awareness. *IEEE Commun. Mag.* 56, 9 (2018), 87–93.
[232] Shulong Wang, Yibin Hou, Fang Gao, and Xinrong Ji. 2016. A novel IoT access architecture for vehicle monitoring system. In *IEEE 3rd World Forum on Internet of Things (WF-IoT'16)*. 639–642.
[233] Wendong Wang, Cheng Feng, Bo Zhang, and Hui Gao. 2019. Environmental monitoring based on fog computing paradigm and internet of things. *IEEE Access* 7 (2019), 127154–127165.
[234] P. D. Wegener. 2018. German standardization roadmap industrie 4.0 version 3. *DIN e* 2018 (2018).
[235] Jiantao Wei, Naiqian Zhang, Ning Wang, Donald Lenhert, Mitchell Neilsen, and Masaaki Mizuno. 2005. Use of the "smart transducer" concept and IEEE 1451 standards in system integration for precision agriculture. *Comput. Electron. Agric.* 48, 3 (2005), 245–255.
[236] Alexander Willner, Christian Diedrich, Raéd Ben Younes, Stephan Hohmann, and Andreas Kraft. 2017. Semantic communication between components for smart factories based on oneM2M. In *22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA'17)*. IEEE, 1–8.
[237] Krzysztof Witkowski. 2017. Internet of things, big data, industry 4.0—Innovative solutions in logistics and supply chains management. *Procedia Eng.* 182 (2017), 763–769.
[238] Huifeng Wu, Junjie Hu, Jiexiang Sun, and Danfeng Sun. 2019. Edge computing in an IoT base station system: Reprogramming and real-time tasks. *Complexity* 2019 (2019). https://www.hindawi.com/journals/complexity/2019/4027638/

[239] Huifeng Wu, Danfeng Sun, Lan Peng, Yuan Yao, Jia Wu, Quan Z. Sheng, and Yi Yan. 2019. Dynamic edge access system in IoT environment. *IEEE Internet Things J.* 7, 4 (2019), 2509–2520.

[240] Huifeng Wu, Yi Yan, Danfeng Sun, and Simon Rene. 2019. A customized real-time compilation for motion control in embedded PLCs. *IEEE Trans. Industr. Inform.* 15, 2 (2019), 812–821.

[241] Huifeng Wu, Yi Yan, Danfeng Sun, and Rene Simon. 2019. VCA protocol-based multilevel flexible architecture on embedded PLCs for visual servo control. *IEEE Trans. Industr. Electron.* 67, 3 (2019), 2450–2459.

[242] Yulei Wu, Hong-Ning Dai, Haozhe Wang, Zehui Xiong, and Song Guo. 2022. A survey of intelligent network slicing management for industrial IoT: Integrated approaches for smart transportation, smart energy, and smart factory. *IEEE Commun. Surv. Tutor.* 24, 2 (2022), 1175–1211.

[243] Jost Wübbeke, Mirjam Meissner, Max J. Zenglein, Jaqueline Ives, and Björn Conrad. 2016. Made in China 2025. *Mercator Instit. China Studies. Pap. China* 2 (2016), 74.

[244] Xiangdong Hao, Fei Li, and Xiaoguang Gao. 2015. Construction of information fusion system based on cloud computing. In *4th International Conference on Computer Science and Network Technology (ICCSNT'15)*. 1461–1465.

[245] Yonggang Xiao, Yanbing Liu, and Tun Li. 2020. Edge computing and blockchain for quick fake news detection in IoV. *Sensors* 20, 16 (2020), 4360.

[246] Xu Xin, Zhang Yan, Hao Yueying, Jiang Yulei, and Geng Mingzhi. 2022. Research of container security reinforcement multi-service APP deployment for new power system on substation. In *4th Asia Energy and Electrical Engineering Symposium (AEEES'22)*. 945–949.

[247] Ying Xiong, Yulin Sun, Li Xing, and Ying Huang. 2018. Extend cloud to edge with KubeEdge. In *IEEE/ACM Symposium on Edge Computing (SEC'18)*. 373–377.

[248] Hansong Xu, Wei Yu, David Griffith, and Nada Golmie. 2018. A survey on industrial internet of things: A cyber-physical systems perspective. *IEEE Access* 6 (2018), 78238–78259.

[249] Yang Xu, Ju Ren, Guojun Wang, Cheng Zhang, Jidian Yang, and Yaoxue Zhang. 2019. A blockchain-based nonrepudiation network computing service scheme for industrial IoT. *IEEE Trans. Industr. Inform.* 15, 6 (2019), 3632–3641.

[250] Jianghui Yan, Jinping Liu, and Fang-Mei Tseng. 2020. An evaluation system based on the self-organizing system framework of smart cities: A case study of smart transportation systems in China. *Technol. Forecast. Social Change* 153 (2020), 119371.

[251] Linfu Yang and Bin Liu. 2019. Temporal data fusion at the edge. In *IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS)*. IEEE, 9–14.

[252] Lina Yao, Quan Z. Sheng, Xue Li, Tao Gu, Mingkui Tan, Xianzhi Wang, Sen Wang, and Wenjie Ruan. 2017. Compressive representation for device-free activity recognition with passive RFID signal strength. *IEEE Trans. Mob. Comput.* 17, 2 (2017), 293–306.

[253] Abbas Yazdinejad, Reza M. Parizi, Ali Dehghantanha, Qi Zhang, and Kim-Kwang Raymond Choo. 2020. An energy-efficient SDN controller architecture for IoT networks with blockchain-based security. *IEEE Trans. Serv. Comput.* 13, 4 (2020), 625–638.

[254] Xun Ye and Seung Ho Hong. 2019. Toward industry 4.0 components: Insights into and implementation of asset administration shells. *IEEE Industr. Electron. Mag.* 13, 1 (Mar. 2019), 13–25.

[255] ChuanTao Yin, Zhang Xiong, Hui Chen, JingYuan Wang, Daven Cooper, and Bertrand David. 2015. A literature survey on smart cities. *Sci. China Inf. Sci.* 58, 10 (2015), 1–18.

[256] Matti Yli-Ojanperä, Seppo Sierla, Nikolaos Papakonstantinou, and Valeriy Vyatkin. 2019. Adapting an agile manufacturing concept to the reference architecture model industry 4.0: A survey and case study. *J. Industr. Inf. Integ.* 15 (Sept. 2019), 147–160.

[257] Xiao Yue, Huiju Wang, Dawei Jin, Mingqiang Li, and Wei Jiang. 2016. Healthcare data gateways: Found healthcare intelligence on blockchain with novel privacy risk control. *J. Med. Syst.* 40, 10 (2016), 218.

[258] Matei Zaharia, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauly, Michael J. Franklin, Scott Shenker, and Ion Stoica. 2012. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. In *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI'12)*. 15–28.

[259] František Zezulka, P. Marcon, Zdenek Bradac, Jakub Arm, T. Benesl, and Ivo Vesely. 2018. Communication systems for industry 4.0 and the IIoT. *IFAC-PapersOnLine* 51, 6 (2018), 150–155.

[260] Daniel Zhang, Nathan Vance, and Dong Wang. 2019. When social sensing meets edge computing: Vision and challenges. In *28th International Conference on Computer Communication and Networks (ICCCN'19)*. IEEE, 1–9.

[261] Lingwen Zhang, Ning Xiao, Wenkao Yang, and Jun Li. 2019. Advanced heterogeneous feature fusion machine learning models and algorithms for improving indoor localization. *Sensors* 19, 1 (2019), 125.

[262] Lili Zhang, Yuxiang Xie, Luan Xidao, and Xin Zhang. 2018. Multi-source heterogeneous data fusion. In *International Conference on Artificial Intelligence and Big Data (ICAIBD'18)*. 47–51.

[263] Qikun Zhang, Yongjiao Li, Ruifang Wang, Lu Liu, Yu-an Tan, and Jingjing Hu. 2021. Data security sharing model based on privacy protection for blockchain-enabled industrial Internet of Things. *Int. J. Intell. Syst.* 36, 1 (2021), 94–111.

[264] Yin Zhang, Yongfeng Qian, Di Wu, M. Shamim Hossain, Ahmed Ghoneim, and Min Chen. 2019. Emotion-aware multimedia systems security. *IEEE Trans. Multim.* 21, 3 (2019), 617–624.

[265] Zheng Zhang, Liang Huang, Renzhong Tang, Tao Peng, Lihang Guo, and Xingwei Xiang. 2020. Industrial blockchain of things: A solution for trustless industrial data sharing and beyond. In *IEEE 16th International Conference on Automation Science and Engineering (CASE'20)*. 1187–1192.

[266] Ma Zhaofeng, Wang Xiaochang, Deepak Kumar Jain, Haneef Khan, Gao Hongmin, and Wang Zhen. 2020. A blockchain-based trusted data management scheme in edge computing. *IEEE Trans. Industr. Inform.* 16, 3 (2020), 2013–2021.

[267] Pai Zheng, Zhiqian Sang, Ray Y. Zhong, Yongkui Liu, Chao Liu, Khamdi Mubarok, Shiqiang Yu, and Xun Xu. 2018. Smart manufacturing systems for industry 4.0: Conceptual framework, scenarios, and future perspectives. *Front. Mechan. Eng.* 13, 2 (2018), 137–150.

[268] Funa Zhou, Po Hu, Shuai Yang, and Chenglin Wen. 2018. A multimodal feature fusion-based deep learning method for online fault diagnosis of rotating machinery. *Sensors* 18, 10 (2018), 3521.

[269] Yuqing Zhou and Wei Xue. 2018. A multisensor fusion method for tool condition monitoring in milling. *Sensors* 18, 11 (2018), 3866.

[270] Tao Zhu, Sahraoui Dhelim, Zhihao Zhou, Shunkun Yang, and Huansheng Ning. 2017. An architecture for aggregating information from distributed data nodes for industrial internet of things. *Comput. Electric. Eng.* 58 (2017), 337–349.

Received 10 May 2022; revised 3 July 2023; accepted 11 July 2023