The Role of Big Data Analytics in Industrial Internet of Things

Muhammad Habib ur Rehman[a], Ibrar Yaqoob[b], Khaled Salah[c], Muhammad Imran[d], Prem Prakash Jayaraman[e], Charith Perera[f]

[a]*National University of Computer and Emerging Sciences, Lahore, Pakistan*
[b]*Department of Computer Science and Engineering, Kyung Hee University, Yongin-si 17104, South Korea*
[c]*Department of Electrical and Computer Engineering, Khalifa University, UAE*
[d]*College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia*
[e]*Department of CS & SE, Swinburne University of Technology, VIC 3122, Australia*
[f]*School of Computer Science and Informatics, Cardiff University, United Kingdom*

Abstract

Big data production in industrial Internet of Things (IIoT) is evident due to the massive deployment of sensors and Internet of Things (IoT) devices. However, big data processing is challenging due to limited computational, networking and storage resources at IoT device-end. Big data analytics (BDA) is expected to provide operational- and customer-level intelligence in IIoT systems. Although numerous studies on IIoT and BDA exist, only a few studies have explored the convergence of the two paradigms. In this study, we investigate the recent BDA technologies, algorithms and techniques that can lead to the development of intelligent IIoT systems. We devise a taxonomy by classifying and categorising the literature on the basis of important parameters (e.g. data sources, analytics tools, analytics techniques, requirements, industrial analytics applications and analytics types). We present the frameworks and case studies of the various enterprises that have benefited from BDA. We also enumerate the considerable opportunities introduced by BDA in IIoT. We identify and discuss the indispensable challenges that remain to be addressed as future research directions as well.

Keywords: Internet of Things, cyber-physical systems, cloud computing, analytics, big data.

1. Introduction

Industrial Internet of Things (IIoT) (also known as Industry 4.0), which was initially conceived as a vision by the German government, is currently attributed as the fourth industrial revolution. The technology ecosystem underpinning IIoT is mainly the integration of cyberphysical systems (CPS) [1], Internet of Things (IoT), cloud computing [2–4], automation (e.g. intelligent robots in product assembly lines) [5], Internet of services [6], wireless technologies, augmented reality [7] and concentric computing [8], amongst others. Advances in such related areas as IoT, big data analytics (BDA), cloud computing and CPS have fuelled the formation of IIoT activities to deliver unprecedented flexibility, precision and efficiency to manufacturing processes [9, 10]. Given this cross-platform integration, IIoT systems need to ensure interoperability, virtualisation, decentralisation, real-time capability, service orientation, modularity and security across all verticals [11]. However, these systems are perceived to have qualities, such as self-awareness, self-prediction, self-comparison, self-configuration, self-maintenance and self-organisation [12].

BDA is a related area that enables IIoT systems to deliver value for data captured from cross-platform integration. BDA refers to the process of collecting, managing, processing, analysing and visualising continuously evolving data in terms of volume, velocity, value, variety and veracity [13]. Big data in IIoT systems arise due to unbounded internal and external activities relevant to customers, business operations, production and machines [14]. BDA processes in IIoT systems manage the collected data using multiple transient and persistent storage systems that provide on-board, in-memory, in-network and large-scale distributed storage facilities across IIoT systems [15, 16]. The granularity of data processing facilities for BDA processes in IIoT systems vary from resource-constrained IoT devices to resourceful large-scale distributed cloud computing systems [17]. Similarly, analytic operations differ in terms of descriptive, prescriptive, predictive and preventive procedures [14]. In addition, BDA processes must ensure real-time knowledge visualisation across multiple IIoT systems. A proper integration of BDA processes into IIoT systems is perceived to maximise value creation to evolve business models for profit maximisation [14, 18].

1.1. Motivation

Although IIoT [19–24] and BDA [13, 25–31] have been widely studied separately, only a few studies including [32] have explored the convergence of the two domains.

Big data production in IIoT is evident due to large-scale deployment of sensing devices and systems in pervasive and ubiquitous industrial networks. Given that the concept of IIoT systems is still evolving, complete integration and implementation of BDA processes in IIoT systems are unavailable yet [32, 33]. Existing surveys on IIoT systems focus on concepts related to adoption of IIoTs [34, 35], the integration of IIoTs and edge cloud computing systems [36], industrial marketplaces for IIoTs [4], big data and virtualisation technologies for IIoT systems [37], technological advancements relevant to CPS in IIoT systems [38], smart manufacturing [39] and big data applications for business operations [40–42]. We introduced the concept of the concentric computing model (CCM) for BDA in IIoT in our previous work [32] whereby we outlined the discussion on different layers of CCM and discussed the relevant research challenges that must be addressed to fully enable CCM for BDA in IIoT. However, to the best of our knowledge, a detailed review on BDA implementation for IIoTs is still lacking in the existing literature. Thus, the current study presents the key operations of BDA for value creation in IIoT systems. On the basis of BDA concepts, this study surveys earlier contributions relevant to data analysis in IIoT systems.

1.2. Contributions

The main contributions of this study are listed as follows.

- We build a case of BDA for IIoT systems whereby the role and entire process of BDA are discussed. The study sets a theoretical ground to understand modern automated data pipelines for enriching intelligence in IIoT systems.
- We investigate existing state-of-the-art research studies on IIoT in terms of BDA. In this context, we categorise and classify the literature by devising a taxonomy.
- We present frameworks and case studies whereby BDA processes are adopted to improve the overall performance of IIoT systems.
- We present several research opportunities, challenges and future technologies to minimise the research gaps between state of the art (i.e. proposed in the literature) and state of the practice (i.e. adopted by industries in practice).

The rest of the paper is organised as follows. Section 2 discusses the key concepts relevant to BDA in IIoT systems, followed by a detailed survey of existing technologies and algorithms in Section 3. Section 4 presents the taxonomy, and Section 5 highlights a few frameworks and relevant case studies. Section 6 presents the opportunities, open challenges and future directions. Section 7 provides the concluding remarks.

2. BDA in IIoT Systems

This section presents a detailed discussion on different aspects of big data adoption in IIoT systems. To this end, several design principles, which should be considered prior to configuring and deploying IIoT systems, are highlighted. The role of BDA and its life cycle is discussed in detail to deliver end-to-end intelligence in IIoT systems.

2.1. Design Principles for IIoT Systems

The designs of IIoT systems involve seven principles [11], as depicted in Fig. 1. Firstly, interoperability must be ensured amongst different technologies, such as CPS, IoT devices and concentric computing systems. Wireless data communication technologies play an unparalleled role to realise an interoperable system. Secondly, virtualisation technologies at all levels must be considered for efficient service provisioning and delivery across IIoT systems. Virtualisation varies in terms of platforms, networks, data, operating systems and applications. Thirdly, decentralisation must be conducted to ensure highly distributed IIoT systems. Decentralisation varies in terms of system-wide data processing and data storage. Fourthly, IIoT systems must provide real-time feedback to all stakeholders. Fifthly, service-orientation must be guaranteed whereby all system functions are implemented in the form of service-oriented architecture (SOA). Sixthly, modular approach must be adopted for system implementation. Lastly, system-wide security must be considered as core principle. The BDA process for IIoT systems must be designed in consideration of the above-mentioned principles.

2.2. Rise of Big Data in IIoT Systems

Big data in IIoT systems emerge from a plethora of technologies. CPS refers to the integration of physical machine components with on-board computations and networking facilities [38, 43]. CPS and IoT devices act as the backbone of IIoT systems and thus generate massive amount of raw data streams, which result in big data [44]. Therefore, real-time analysis of these data can improve machine health and lead to defect-free product manufacturing [1, 34, 45].

IoT devices in IIoT systems refer to devices that can remotely sense and actuate in industrial environments [46]. IoT devices either work as stand-alone devices that roam around industrial environments or are attached with existing CPS to perform certain predefined actions [47]. The on-board sensing facilities in IoT devices lead the generation of big data, which may become

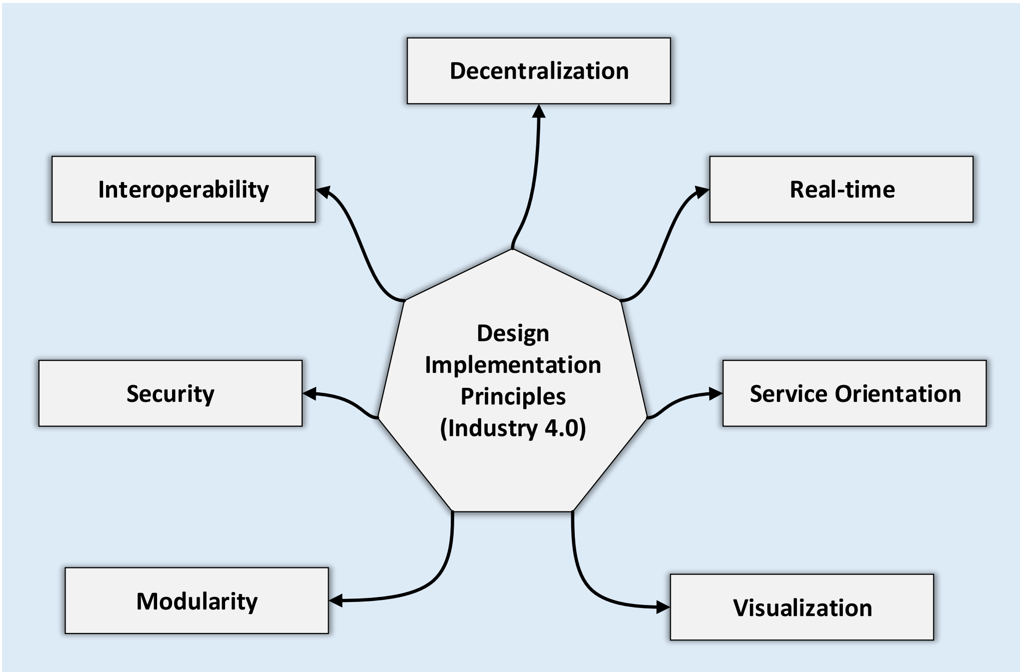

Figure 1: Seven design implementation principles for Industry 4.0 systems.

useful for value creation in enterprises. The integration of CPS and IoT devices results in massive back-end cloud service utilisation for the execution of BDA processes [48]. To achieve massively customised production, the number of cloud services can be grown immensely. Thus, BDA can facilitate in-service selection, service orchestration and real-time service provisioning [49].

2.3. Concentric Computing Model for BDA in IIoT

Recent evolution in sensing and computing technologies has opened new avenues for big data processing. Concentric computing refers to the large-scale highly distributed computing systems based on a wide range of devices and computing facilities in different form factors [8]. Concentric computing offers big data processing at sensors levels, endpoints in IIoT systems, edge servers, and centralised and decentralised cloud computing systems, as illustrated in Fig. 2 [14, 36, 50]. Despite their small size and limited computational power, sensors and IoT devices have the ability to filter and reduce raw data streams by using on-board smart data reduction strategies [51]. However, edge servers at gateways and centralised computing clusters have the ability to distribute the computing load for BDA applications [52, 53]. Multistage execution, automating, and management of BDA processes (i.e., data engineering, data preparation and data analytics) are necessary in concentric computing environments (such as sensors and wearable devices as endpoints, IoT devices, edge servers, and cloud computing servers) [54].

2.4. Big Data Analytics for Delivering Intelligence in IIoT Systems

BDA processes are executed as a result of multistage highly interdependent application components (Fig. 3). These components are categorised as follows.

2.4.1. Data Engineering

Data engineers build computing and storage infrastructure to ingest [55], clean [56], conform [57], shape [58] and transform [59] data. IIoT systems produce and ingest big data from inbound enterprise operations and outbound customer activities. The raw data at the earliest stage need further processing to improve the quality and establish the relevance with IIoT applications. Therefore, data wrangling and cleaning methodologies help select relevant datasets in case of historical data or data streams in case of streaming data. Data conformity procedures are applied to ensure relevant, correctly collected big data. Data shaping and transformation

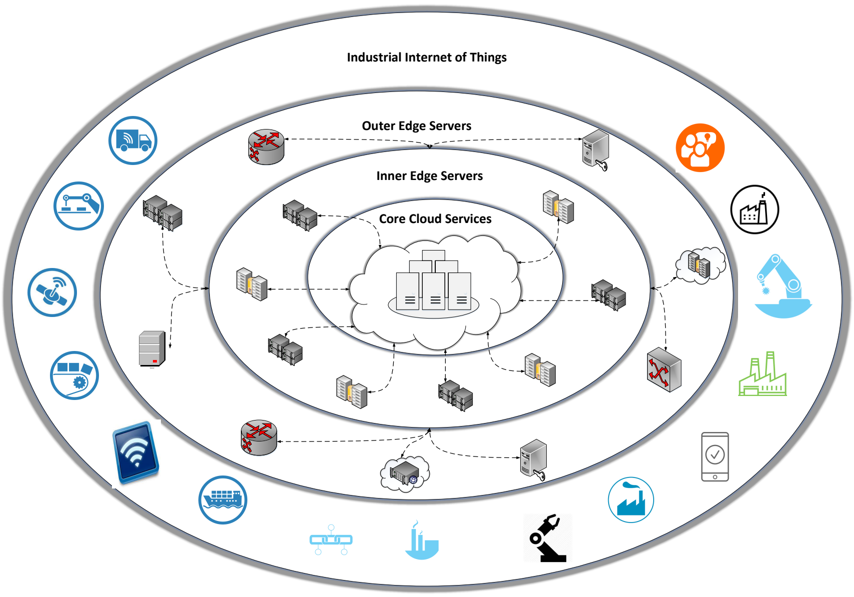

Figure 2: Industrial IoTs and Multilayer Computing Resources

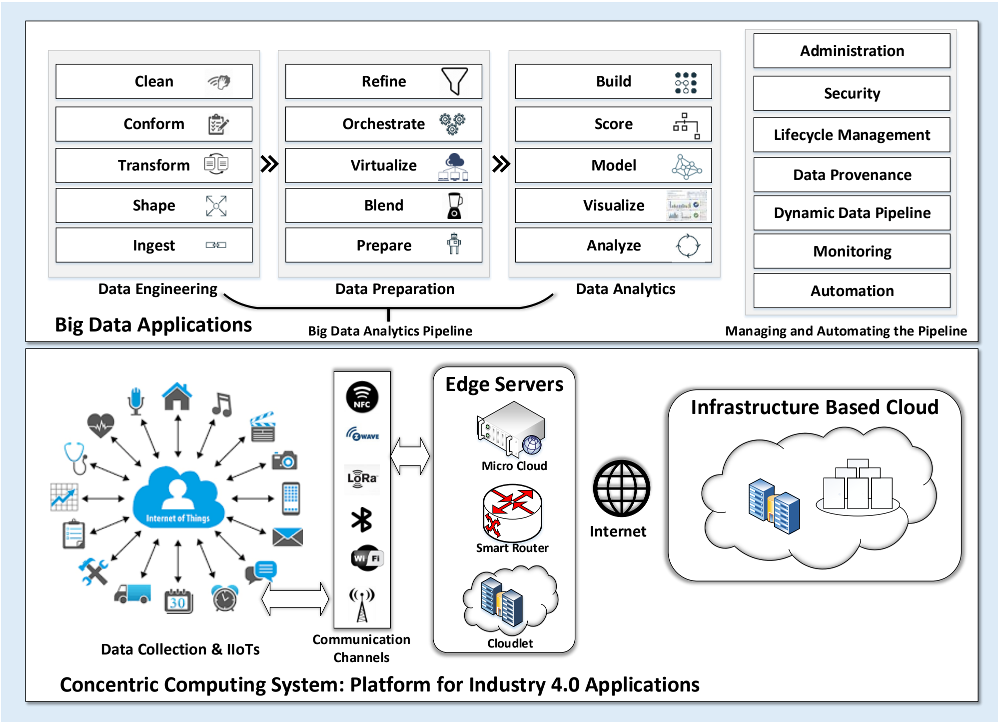

Figure 3: Multistage execution, automating, and management of BDA processes (i.e., data engineering, data preparation and data analytics) in concentric computing environments (such as sensors and wearable devices as endpoints, IoT devices, edge servers, and cloud computing servers) [32].

methodologies help improve data quality by reducing the number of attributes and converting data formats for uniform data processing.

2.4.2. *Data Preparation*

Big data emerge in raw form with large volume and enormous speed, and data scientists spend 70% – 80% of their time in data preparation activities [60]. Big data are refined using statistical methods to handle unstructured, unbalanced and nonstandardised data points efficiently [61]. In addition, data refinement helps summarise voluminous data to reduce overall complexity. As a result, the spatiotemporal attributes of big data in IIoT systems vary. Ultimately, data locality is necessary to reduce in-network traffic and latency in big-data applications [61]. Location-aware highly virtualised data infrastructure can address these issues. However, data blending, which is the process of combining data from multiple sources, becomes complex. Accordingly, further involvement by data scientists [62] is required to perform data cleaning and noise removal [63]. Detection methods for outliers and anomalies are also needed to prepare big data for further analysis [64, 65].

2.4.3. Data Analytics

The analytic processes in IIoT systems are executed in multiple phases [66]. Data scientists generate learning models from high-quality well-prepared data. After the model is developed, model scoring operations are performed by giving sample datasets and finding and ranking the attributes in datasets/data streams. The correctly tuned models are deployed in production environments to find the knowledge patterns from future data.

2.4.4. Managing and Automating the Data Pipeline

Although existing literature still lacks the concept of automated data pipelines in IIoT systems, BDA processes are executed as a sequence of operations during data engineering, preparation and analytics. Therefore, a holistic approach is needed to execute and administer BDA processes across all layers of concentric computing systems. Life cycle management is needed for full process execution from raw data acquisition to knowledge visualisation and actuation. Data provenance, that is, designating ownership of data to different stack holders, also needs serious attention to ensure system-wide control on data [67]. The continuous evolution in data streams results in knowledge shift that enforces data pipelines to adaptively reconfigure analytic processes. The data pipelines need to be continuously monitored for change detection, and the entire BDA process needs to be re-executed to produce high-quality results [68]. In security perspective, the cross-platform execution of BDA processes demands secure operations at IoT device, CPS and big data levels [69].

3. Technologies and Algorithms for BDA in IIoT systems

A common example of IIoT systems is the concept of a smart factory system (SFS) [36]. The key attributes of SFS and its subsystems are self-awareness, self-organisation, self-maintenance, self-prediction, self-configuration and self-comparison [12]. This section presents the review of early studies that presented BDA in the context of SFS and IIoT systems [12] in consideration of the aforementioned autonomy related attributes (Table 1).

3.1. Mass Product Customization towards IIoT Lean Manufacturing

Although the main objective of IIoT systems is to maximise production considering massive customisation in accordance with customer requirements, the existing literature still lacks an end-to-end predictive analytics framework. Computational intelligence-based methods, such as self-organising map (SOM) algorithms, are used to optimise big data for feeding in the production systems and enable massively customised product manufacturing [70]. The neural network-based SOM algorithm effectively enables smart production cycle in SFS. The cycle is based on a close loop within a sequence of operations, including smart design, manufacturing, production and services whereby feedback is collected after each cycle and subsequent operations at each stage are improved. Clustering-based big data optimisation is another approach whereby k-means clustering algorithms are used to cluster the attributes from customer data. The produced clusters are used to intelligently improve the design process in the product life cycle [71]. Another alternate for massive product customisation is the adoption of cloud-based manufacturing systems whereby big data integration is performed in cloud computing systems [72]. However, the resultant big data are integrated from multiple sources, such as social media data streams relevant to customer behaviour and IIoT data streams from manufacturing systems. This type of cloud-based manufacturing benefits from open innovation and cross-continent physically isolated product manufacturing.

3.2. Industrial Time Series Modeling

The achievement of zero-defect in SFS is a major challenge. In SFS, all manufacturing components are perceived to be highly connected to ensure high-quality production. The term zero-defect refers to ensuring high-quality production during the execution of a complete manufacturing process [45]. To this end, industrial time series modelling ensures the proper monitoring of all manufacturing components during operations. However, data collection from multiple components results in high-dimensional data streams. The neo-fuzzy neuron (NFN) time series modelling method is adopted by IIoT systems. NFN can collaboratively connect the input data streams with the final outputs. NFN benefits from the convergence of input data, which results in decreased data streams and thus less iteration for learning model generation [73].

3.3. Intelligent Shop Floor Monitoring

The term physical Internet (PI) refers to the integration of cloud manufacturing with wireless and networking technologies. PI in IIoT systems provides the backbone to IIoTs and smart manufacturing object tracking systems based on radio-frequency identification. These smart manufacturing objects represent different forms of products during manufacturing after each process [47]. However, IIoT systems need to track these smart objects during production to ensure that analytic processes provide intelligent shop monitoring. Researchers have proposed a BDA-based approach for the trajectory clustering of moving objects in shop floors. Although initial findings have been previously presented, a component-based BDA architecture is still necessary to develop highly optimised and intelligent smart object tracking systems for shop floor monitoring [47]. Performance analysis and exception diagnosis model have been proposed and tested using Petri nets and decision tree algorithms [47]. The model shows feasibility, and its real implementation in IIoT systems may help correctly quantify the results.

3.4. Industrial Microgrids

Massive data production in IIoT systems is evident due to feature-rich sensory and large-scale deployment of IIoTs in SFS [74]. Therefore, manufacturing and environmental data, along with energy consumption data, can lead towards optimised energy utilisation in SFS. The application of BDA processes on these data silos can help improve planning, managing and utilising energy. Researchers have proposed BDA analytics methods for industrial-level microgrid planning in SFS. However, quantifiable studies that can lead towards efficient microgrid planning in IIoT systems are still required [74].

3.5. Monitoring Machine Health

Prognostic health monitoring (PHM) helps find the machine behaviour for value creation during mechanical operations and facilitate machine data collection and management for the early diagnoses and prediction of machine faults. Several studies have performed analysis of PHM data [75–77]. In accordance with multiple International Standards Organisation and International Electrotechnical Commission and Society of Automotive Engineering standards, the authors of [76] analysed ontological models developed from PHM data. These ontological models represent the hierarchical and semantic relationships amongst different machine components. The remaining useful life of machine components, faults, errors and failures during machine operations has also been explored. Studies have also presented dependency and failure mode analyses of different machine components. The analysis of PHM data helps plan and schedule machinery maintenance activities, thereby supporting in finding maintainable machine components before total failure. However, finding the relationship amongst different attributes and the failure impact of understudied machine components on other components in large-scale manufacturing environments is a challenging task [77].

3.6. Intelligent Predictive and Preventive Maintenance

Predictive and preventive maintenance are the key requirements of large-scale IIoT systems [11]. The BDA process can help in off-line prediction (*i.e.,* performing prediction on the basis of historical data) and online maintenance (*i.e.,* maintaining machines without shutting down the manufacturing units). Researchers have integrated Hadoop and Storm technologies for big data processing and used neural network-based methods for prediction [78]. The concept of adopting BDA for intelligent predictive maintenance is novel. However, new avenues need to be explored to fully realise a real-time prediction system.

4. Taxonomy

Figure 4 presents the taxonomy that is devised on the basis of data sources, analytics tools, analytics techniques, requirements, industrial analytics applications and analytics types.

4.1. Data Sources

In an industrial environment, numerous sources of data production, such as sensors, enterprise resource planning (ERP) systems, manufacturing execution systems (MES), supervisory control and data acquisition (SCADA) systems, customer relationship management (CRM) systems and machine/IoT devices. ERP systems enable organisations to employ a system that is composed of multiple integrated applications for managing business needs and automating many back-office functions related to technology, services and human resources. MES helps keep the track record of all manufacturing information in real time and receive up-to-date data from robots, machine and IoT devices [79]. SCADA systems are used to monitor and control a plant or equipment in industries (*e.g.* telecommunications, water and waste control, energy, oil and gas refining and transportation). CRM systems are commonly used to manage a businesscustomer relationship. Machines and IoT devices are also deployed in industries to perform specific tasks, which generate an enormous amount of data on a daily basis. Applying analytics solutions to the collected data through all the above-mentioned systems, machines and IoT devices can extract valuable information that can help in decision-making purposes.

4.2. *Analytics Tools*

Several analytics tools are required to gain insights into a large amount of industrial data. These tools include analytics software, algorithm repository, visualisation tools, modelling tools and online analytics packages. Analytics software helps make predictions about unknown events. An algorithm repository is a crowd-sourced repository of algorithms that is designed by analysts using a common set of languages and a common interface. Visualisation tools help present data in advanced formats (e.g. infographics, dials and gauges, geographic maps, sparklines, heat maps and detailed bar, pie and fever charts). Modelling tools are used to define and analyse data requirements for supporting business processes within the scope of corresponding information systems in industries. Online analytics packages help keep track of and analyse data about web traffic.

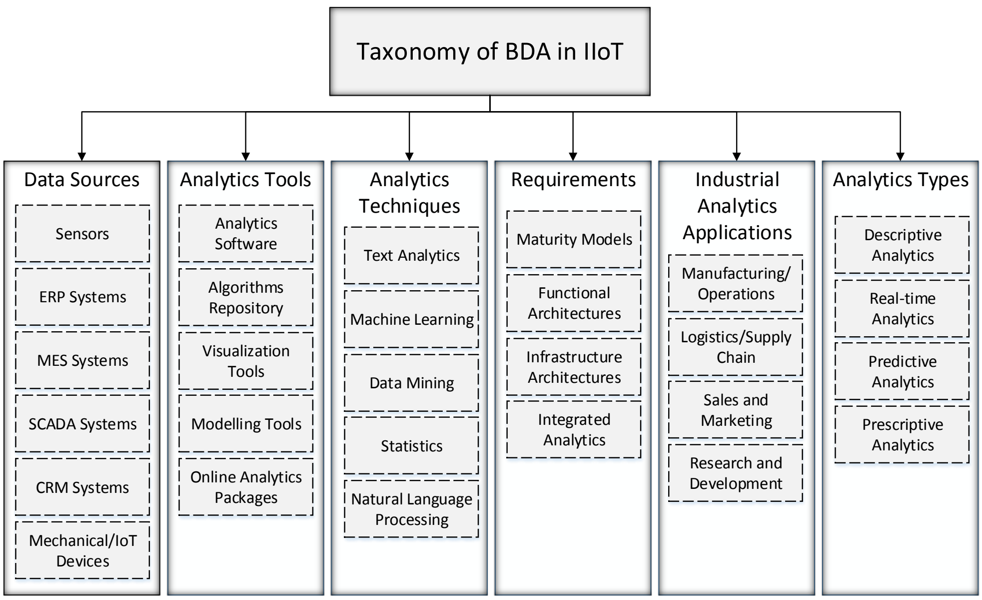

Figure 4: Taxonomy of BDA in IIoT

Table 1: BDA Implementations in IIoT Systems

Ref.	Problem(s)	Objective(s)	Analytic Component(s)	Mode	Strengths	Limitations	Potential Solutions
[71]	Finding accurate customers' attributes for mass customization.	Self-prediction	Genetic Algorithm k-means clustering	Historical data	Smart product development Rapid response to customer needs	Needs to be implemented for real-time, Lacks large-scale validation in BDA environments	Deep Learning for BDA
[70]	Finding accurate customers' attributes for mass customization.	Self-prediction	Self-organizing map	Historical data	Smart product development Rapid response to customer needs	Needs to be implemented for real-time, Lacks large-scale validation in BDA	Re-enforcement learning algorithms
[72]	Enabling product customization and personalisation	Self-configuration Self-organization	Highlighted, but no real implementation discussed	Streaming data	An end-to-end model for massive production and personalisation	No real implementation	Use-case implementation
[73]	Achieving zero-defect problem	Self-configuration	Neo-Fuzzy Neuron	Batch data	Performs industrial process monitoring and modelling	Accuracy needs to be implemented	Using alternate ML algorithms
[47]	implementing Physical Internet concept in manufacturing shop floors	Self-prediction	Decision trees	Batch data	The implementation results in better prediction rate	Performance values for different workers needs to be well-defined to control the rate of overestimation	Using alternate ML algorithms
[74]	Developing a proactive and sustainable microgrid	Self-prediction	A generic framework for knowledge discovery	Batch Data	An end-to-end approach for microgrid data analysis	Efforts are needed to explore analytics for full value chain level knowledge discovery in industrial microgrids	BDA Platform for full value chain Analytics
[78]	Active preventive maintenance	Self-maintenance	Neural Networks	Batch data	Real-time active maintenance	Need to be investigated with real-time streaming data	Real-time BDA platform

4.3. Analytics Techniques

Various analytics techniques that can help obtain value from big industrial data are available, thereby leading to making fast and better decisions. These analytics techniques include text analytics, machine learning, data mining and statistical and natural language processing (NLP) techniques. Text analytics helps derive high-quality information by unveiling patterns and trends using statistical pattern learning. Machine learning techniques enable industrial devices and machines to enter into a self-learning mode without being explicitly programmed. Data mining solutions enable enterprises to transform raw data into knowledge. Statistical tools help collect, summarise, analyse and interpret large amounts of industrial data, which lead to knowledge discovery. In an industrial environment, NLP tools are used to extract and analyse unstructured data.

4.4. Requirements

Certain requirements should be incorporated whilst developing new analytics systems for IIoT. These requirements include maturity models, functional architecture, infrastructure architecture and integrated analysis. Maturity models help measure and monitor the capabilities of analytics systems. They also help measure the effort required to complete a specific development stage. In summary, these models help monitor the health of an organisations big data programs. Functional architecture is an architectural model that helps identify the functions of analytics systems and their interactions. In addition, it defines how system functions work together to perform a specified system mission. In an industrial environment, analytics systems must be developed such that they can handle an enormous amount of data in real time. In this context, big data infrastructure requires experienced scientists to design the infrastructure from existing equipment in an industrial paradigm. One of the key requirements for analytics systems is that they should support the integrated analysis of multiple types of industrial IoT data.

4.5. Industrial Analytics Applications

Typical industrial analytics applications across the industrial value chain are as follows: manufacturing/operations, logistics/supply chain, marketing/sales and research and development. The use of predictive analytics in manufacturing can lead to rescheduling a maintenance plan prior to machine failure by considering past machine performance history. Moreover, it can help develop decision support systems for industrial processes. The appropriate use of analytics can play an important role in the logistics/supply chain (*e.g.* condition monitoring, supply chain optimisation, fleet management and strategic supplier management). Analytics can help identify failing parts during product usage through sensor readings and gradually improve product characteristics (research and development). In the marketing field, analytics tools enable enterprises to predict and enhance future sales (*e.g.* help in determining seasonal trends that can lead to developing an adaptive marketing strategy).

4.6. Analytics Types

Analytics has four types: descriptive, real-time, predictive and prescriptive analytics. Descriptive analytics helps gain insights into historical data (*e.g.* number of defective items in the past and the reason for the defects). Meanwhile, real-time analytics enables enterprises to become aware of current situations (*e.g.*, current status and location of a product and detection of a faulty machine). By contrast, predictive analytics helps identify potential issues that can occur in advance by using statistical and machine learning techniques (*e.g.* expected inventory levels, anticipated demand levels, and prediction of equipment failure). Lastly, prescriptive analytics provides advice or suggestion on the best possible action that an end user should take (*e.g.* whether a machine is receiving the right raw materials in the correct amount).

5. Frameworks and Case Studies

Value creation is a major sustainability factor in modern enterprises whereby BDA processes are becoming the primary driver in creating values for customers and enterprises [80]. IIoT systems are no exception. BDA processes can facilitate the amalgamation of customer and enterprise data to ensure massively customised production with zero defects. IIoT systems essentially integrate historical and real-time stored and streaming data at various levels. This multisource data integration leads to highly effective designs for new business models. Enterprises focus on different aspects of industry-wide value creation mechanisms, such as defining value propositions, value capturing mechanisms, value networks and value communication strategies for internal and external stakeholders [18]. Ideally, BDA processes can facilitate enterprise-level value creation whereby inbound intelligence is obtained by creating value for internal enterprise operations. Alternatively, outbound intelligence leads towards value creation for customers. Despite these opportunities, unlocking the perceived value from BDA technologies is challenging. The existing literature presents only a few such frameworks and use cases as follows.

5.1. SnappyData

SnappyData is an open-source BDA framework that integrates Apaches Spark and GemFire technologies [81]. Apaches Spark is adopted for big data processing, whereas GemFire facilitates highly scalable in-memory transactional data storage. The strength of SnappyData is its unified BDA engine that facilitates the performance of different types of analytical operation, such as online transaction processing, online analytical processing and streaming the data analytics of operational data. Despite its high performance, SnappyData still underperforms in cases with highly streaming data, which causes a bottleneck in real-time interactive visualisation performance.

5.2. Ipanera

Soilless food production systems, such as Ipanera, are being aligned with IIoT systems [82]. Ipanera continuously monitors water level and fertilizer quality in a field and generates insights for self-configuration. Although researchers have presented the concept, the Ipanera architecture involves multiple layers of physical devices and systems. It includes sensor nodes at the end points that actively collect data streams and transfer them to nearby IIoT clusters. These clusters are responsible for end point management, communication and configuration in a field. In addition, IIoT clusters provide feedback to end points to reconfigure their data collection behaviour. IIoT clusters transfer data streams to distributed analytics servers that run Apaches Hadoop [83], MapReduce [84] and Spark [85] technologies for data processing and BDA. Ipanera provides support for streaming analytics that is used to trigger alerts for end points in case a new event is detected. Persistent storage and on-the-air configurations are two innovative features of the Ipanera architecture. This architecture is currently under development; hence, the complete design of the proposed architecture is still unavailable.

5.3. Fault Detection Classification

Large-scale distributed cyber manufacturing systems are based on multiple interconnected but geographically dispersed manufacturing units [86]. The fault detection and classification (FDC) framework finds manufacturing faults in products. The core of the FDC architecture is the integration of IoT devices into CPS and cloud computing technologies. IoT devices in production facilities continuously collect and analyse data streams to detect various signals that are transferred to back-end cloud servers. These cloud servers execute BDA processes to detect and classify faulty products using deep belief networks based on deep learning methods [87, 88]. FDC was analysed by deploying it in a car headlight manufacturing unit that produced reliable results.

5.4. BDA Architecture for Cleaner Production

The term cleaner production refers to ensuring reduced environmental impacts during the execution of the entire product life cycle. It is based on three phases [89]. The first phase is about product design and manufacturing. The second phase involves product use, service provisioning and maintenance. The third phase is concerned with product remanufacturing, reuse and recycling. Considering the importance of such clean technologies, researchers have proposed a four-stage BDA architecture. In the first stage, the architecture considers value creation objectives during a products life cycle, such as improving product designs and ensuring energy efficiency, proactive maintenance and environmental efficiency. In the second stage, big data acquisition and integration are performed using IoT devices. In the third stage, big data are processed using Apaches Hadoop and Storm technologies. Finally, BDA processes are executed in the fourth stage whereby the architecture provides clustering, classification, association rule mining and prediction-related algorithms. The proposed architecture was evaluated and tested on an axial compressor manufacturing unit. The annual reports of the production unit show that the proposed architecture realises all the value creation objectives for cleaner production.

5.5. Smart Maintenance Initiative: Railway Case Study

Apart from SFS, Japan is attempting to upgrade its railway system to a new level by adopting IIoT systems for the smart maintenance of railway tracks [90]. To achieve its smart maintenance vision, Japans railway is adopting IIoT, BDA and automation technologies. The smart maintenance vision will provide a solution to four challenges: 1) ensuring condition-based maintenance, 2) providing work support through artificial intelligence (AI), 3) managing railway assets and 4) performing database integration. The progress details of Japans railway towards this vision are available in this report [90] for interested readers.

6. Opportunities, Research Challenges, and Future Technologies

Considering the vision of IIoT systems, BDA will evidently help enterprises in the value creation process. BDA processes will maximise operational efficiency, reduce product development cost, ensure massively customised production and streamline the supply chain management. However, this review shows that the existing literature is considerably lagging behind this vision. Table 2 presents the summary of research challenges and their perceived solutions to fully adopt IIoT systems in BDA.

6.1. Opportunities

The adoption of BDA processes in IIoT systems results in multidimensional research opportunities.

6.1.1. Automation and AI

The enrichment of intelligent features can lead towards highly optimised and automated industrial processes [91, 92]. Therefore, AI will be the core component of big data optimisation and analytics, which will result in highly efficient industrial processes [93]. Future IIoT systems will integrate and ingest big data from various online and off-line and inbound and outbound operations. The integration of customer and enterprise data will result in high-dimensional, multi-million variable datasets. AI methods will help optimise and analyse such big datasets [94, 95].

6.1.2. Human Machine Interaction

Wearable computing and augmented reality technologies are leading towards new human-machine interaction models and interfaces [96, 97]. The enrichment of such interaction models with real-time knowledge patterns from big data systems will result in highly productive and rich user interfaces. In addition, robotics technologies (for physical and virtual robots) will be widely adopted by future IIoT systems. Therefore, BDA processes will enrich intelligence to produce highly autonomous and self-sustaining non-obtrusive systems.

6.1.3. Cybersecurity, Privacy, and Ethics

Cybersecurity will become an essential requirement due to connected intelligence in IIoT systems. BDA processes will help provide real-time cyber threat intelligence by analysing security attacks, privacy leaks, unauthorised data access and unethical data collection [98]. In addition, BDA processes will help analyse network and information security-related enterprise data to find anomalies, outliers, threats, attacks and vulnerabilities across IIoT systems [99].

6.1.4. *Universal Standards*

The adoption of BDA processes is still in its initial stage; thus, existing systems may not be compliant with universal standards across all or multiple industries [2, 100]. New universal standards are required to define the type of big data that the industries can collect from customers, determine how data should be secured, preserved and shared and identify the stakeholders who will benefit from the data. In addition, standards must also ensure the perceived benefits to customers in exchange for their personal data. These universal standards will help address ethical issues in big data systems and create value for customers by providing personalised products and services.

6.1.5. *Protocols for Interoperability*

Practically, multiple industries are involved in the entire processfrom customer data acquisition to finished product/service and supply chain management [2]. Interoperability is a major consideration among different industries; however, new protocols are required to realise fully interoperable IIoT systems. These protocols can lead towards value creation for enterprises, although a few questions must be addressed, such as what are the interoperability parameters, how will BDA processes be executed in cross-industry systems and how will heterogeneity in data, computing technologies and industrial production systems be handled. A well-defined interoperability protocol can help answer these questions.

6.1.6. *End-to-end Industrial Analytics*

Big data in IIoT systems evolves from multiple inbound and outbound data sources, such as customer data and operational data from finance, marketing, human resources, IoT devices, CPS and manufacturing systems [101]. Nevertheless, existing systems manage all these data sources separately to execute BDA processes. An opportunity exists to develop an end-to-end industrial analytics pipeline that can handle big data from various data sources in parallel and find highly correlated knowledge patterns that emerge across entire IIoT systems [102].

6.1.7. *Precision Manufacturing*

BDA processes can help enrich precision manufacturing systems [103]. The classification and categorisation of customers needs and behaviour-related data can lead towards innovative product designs. Enterprises will be able to offer the right products and services to the right customers. Precision manufacturing will considerably help in equal value creation for customers and enterprises. Early examples of precision manufacturing systems are available in the healthcare industry [104]. However, these systems should be integrated into IIoT systems [103].

6.2. Research Challenges and Future Technologies

Considering the opportunities, research efforts are required to improve the entire technology ecosystem for IIoT systems.

6.2.1. Big Data Process Integration into IIoT Systems

Ideally, IIoT systems should execute real-time highly interactive big data applications. In practice, however, considerable effort is required for planning, creating, deploying, maintaining and continuously improving domain-specific big data processes for each industry. Future BDA processes should be able to provide real-time knowledge patterns and industry-wide intelligence through single dashboard applications. In this regard, all legacy and state-of-the art data sources should be vertically aligned such that enterprises can easily analyse and correlate different industrial processes and operations.

6.2.2. Orchestrating BDA Applications Using Concentric Computing

Concentric computing systems provide computational and storage support through different devices and systems [8]. Thus, massive heterogeneity should be addressed in terms of processing capabilities, in-memory and disk-based storage systems, battery-powered and fully powered devices and systems and multiple communication channels with varying bandwidth capacities [17]. Big data applications on top of concentric computing systems should be designed by considering efficiency objectives in terms of storage, in-network data movement, energy consumption, privacy, security and real-time knowledge availability [105, 106]. In this regard, priority should be given to devices and systems near data sources. This approach can help maximise value creation for enterprises in terms of operating cost for big data systems. Given that maximum data collection, filtration and processing are performed before data arrive in cloud computing systems, the operational costs for data storage and cloud service utilisation will therefore be minimised [80]. Another benefit of concentric computing systems is their ability to ensure real-time or near real-time intelligence near end points, IoT devices and other data sources in IIoT systems [36].

6.2.3. *Emerging and Complimentary Technologies for IIoT Systems*

On the one hand, BDA adoption is increasing in IIoT systems. On the other hand, IIoT systems should address massive heterogeneity without compromising overall operational efficiency due to emerging, complementary technologies, such as IoT. Considering this condition, a few technologies will become integral parts of future IIoT systems.

Virtualisation is the essence of distributed systems, such as cloud computing systems and concentric computing systems. Virtualisation is traditionally performed at multiple levels, such as operating systems, networks, storage, applications and hardware. Operating system-level virtualisation is the most common whereby operating system kernels and functions are virtualised as virtual machines (VMs). However, the mobility of IoT devices requires continuous VM migration among different computer servers [2, 107]. Containerisation is the emerging technology that is gradually replacing VMs by sharing a single kernel among different applications on the same type of operating systems. Containerisation technologies offers more secure and faster processing; hence, they have become highly beneficial for addressing timeliness and latency issues when BDA processes are executed using VMs [108].

Large enterprises traditionally adopt highly coupled SOAs, which are difficult to test and result in high maintenance cost. Microservices are emerging alternatives to SOAs whereby highly scalable and loosely coupled cloud services are orchestrated [109]. The microservice architecture can be adopted best for BDA processes because these processes should be executed across multiple platforms and devices in IIoT systems [110]. The details of microservice architectures implementation are available in [111] for interested readers.

The multipoint, multisite and high-dimensional data production in IIoT systems results in complex big datasets. Graph and network theories can help reduce this massive complexity [112]. Graph data structures and big graph analytics methods can be adopted to separate, map and analyse big data in different graph formats. The adoption of big graph analytics can lead towards efficient and highly optimised execution of BDA processes across IIoT systems.

Table 2: Summary of Research Challenges and their Perceived Solutions

Type	Issues	Causes	Solutions
Cybersecurity	- Internal Attacks - External Attacks	- Security Vulnerabilities - Openness of Systems	- Intelligent Monitoring Tools Needed - Deployment of End-to-End Security Models is Essential - System-wide Forensic Analysis should be performed periodically
Privacy	- Identity Breaches - Personal Data Theft - Business Data Leakage	- Bad Security Models - Absence of Standard Operating Procedures - Weak Data and Information Sharing Policies	- Using Data Anonymisation Protocols - Privacy preserving interaction models for users, devices, and systems
Big Data Processing	- Bad Data Integration - Missing Data Streams - High Latency	Heterogeneous Data Sources - Mobility and Connectivity Issues - Data overloading and Bandwidth Limitations	- Intelligent Real time Data Fusion - Device-centric big data processing architectures - Concentric Computing Models
Standardization	- Difficulty in Interoperability and and System Integration	- Absence of Global Standardization Body	- Developing Local, Regional, Industry-specific, and Global Standards
Connectivity and Communication	- Bad and Inaccurate Data Transfer - Data Loss - High Latency	- High Mobility - Large Data Streams - Congestion	- Need to create always-on, ultra-high available and reliable communication protocols
Scalability	- Resource Discovery - Data offloading - Data Management	- Low Processing Power at device-end - Massive Data Production - Realtime Actuation	- Near-device data processing, In-memory Data Processing, Edge Computing
System management	- Difficult to deploy, configure, monitor, and control large scale IIoT networks	- Cloud-centric	- Device-centric
Efficiency	- High Energy Utilization - Resources-constraints - Device-overloading	- Always-on IIoT Devices and Systems - Massive and Continuous Data Generation and Device Operations - On-device Data Management and Analytics	- Enabling Energy, memory, and computation-efficient algorithms and processes for big data processing, management and analytics in IIOTs

Emerging technologies, such as fog computing and blockchain, can play a pivotal role in BDA for IIoT [113]. Fog computing has been widely used in IoT devices [114], particularly those for IIoT and smart manufacturing, for localised and timely data processing and storage, and primarily to offset long delays that can be incurred in a cloud environment [115]. Blockchain is the underlying technology for bitcoins; however, it has been foreseen as a distributed ledger that can provide decentralised storage for data generated by IoT devices. Data are stored in a blockchain ledger with high integrity, authenticity, resiliency and trust [116]. All transactions are cryptographically signed by IIoT devices and validated in a decentralised manner without an intermediary. The data origin is validated before being recorded on the ledger. Moreover, blockchain smart contracts can be used to provide decentralised authentication, management and control access to data generated by IIoT devices. Smart contracts are basically codes that are executed by all blockchain miners, and the execution outcome is verified and agreed upon by all mining nodes. Furthermore, given the limited computing, networking and storage capacities of IIoT devices, fog nodes are envisioned to be equipped with cloud and blockchain interfaces in the future to communicate and interface with the cloud environment and the blockchain network [116].

7. Conclusions

The vision of Industry 4.0 to connect manufacturing systems with distributors and consumers can only be achieved by adopting IIoT and BDA processes as core components for value creation. This paper discusses the rise of big data in IIoT systems and presents a detailed survey of related technologies, algorithms, frameworks and case studies. A detailed taxonomy is provided to classify the key concepts in this important research area. Several indispensable frameworks and case studies are outlined and discussed. Furthermore, we present a detailed discussion of future opportunities, technologies and research challenges. We conclude that the adoption of BDA in IIoT systems is still in its early stage. Research on complementary components of IIoT systems, such as IoT devices, augmented reality and CPS, is also in its infancy. Current BDA systems provide generic frameworks for data engineering, preparation and analysis. However, considerable effort is required to alter existing BDA processes to meet the demands of IIoT systems. Future research should be conducted to devise new standards for interoperability among cross-Industry 4.0 BDA platforms and to provide capability for end-to-end reliable application processing by considering the anatomy of concentric computing systems.

Acknowledgement

Imran's work is supported by the Deanship of Scientific Research at King Saud University through Research group No. (RG # 1435-051).

References

[1] I. Akkaya, P. Derler, S. Emoto, E. A. Lee, Systems engineering for industrial cyber–physical systems using aspects, Proceedings of the IEEE 104 (2016) 997–1012.

[2] M. Wollschlaeger, T. Sauter, J. Jasperneite, The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0, IEEE Industrial Electronics Magazine 11 (2017) 17–27.

[3] C. Perera, S. Y. Wakenshaw, T. Baarslag, H. Haddadi, A. K. Bandara, R. Mortier, A. Crabtree, I. C. Ng, D. McAuley, J. Crowcroft, Valorising the iot databox: creating value for everyone, Transactions on Emerging Telecommunications Technologies 28 (2017).

[4] C. Perera, C. H. Liu, S. Jayawardena, M. Chen, A survey on internet of things from industrial market perspective, IEEE Access 2 (2014) 1660–1679.

[5] M. R. Pedersen, L. Nalpantidis, R. S. Andersen, C. Schou, S. Bøgh, V. Krüger, O. Madsen, Robot skills for manufacturing: From concept to industrial deployment, Robotics and Computer-Integrated Manufacturing 37 (2016) 282–291.

[6] L. Liu, Services computing: From cloud services, mobile services to internet of services, IEEE Transactions on Services Computing 9 (2016) 661–663.

[7] S. Kim, M. A. Nussbaum, J. L. Gabbard, Augmented reality ?smart glasses? in the workplace: Industry perspectives and challenges for worker safety and health, IIE transactions on occupational ergonomics and human factors 4 (2016) 253–258.

[8] P. Howard, Concentric analytics for iot, Online, 2017.

[9] I. Yaqoob, E. Ahmed, I. A. T. Hashem, A. I. A. Ahmed, A. Gani, M. Imran, M. Guizani, Internet of things architecture: Recent advances, taxonomy, requirements, and open challenges, IEEE wireless communications 24 (2017) 10–16.

[10] E. Ahmed, I. Yaqoob, A. Gani, M. Imran, M. Guizani, Internet-of-things-based smart environments: state of the art, taxonomy, and open research challenges, IEEE Wireless Communications 23 (2016) 10–16.

[11] K. Wang, Intelligent predictive maintenance (ipdm) system–industry 4.0 scenario, WIT Transactions on Engineering Sciences 113 (2016) 259–268.

[12] J. Lee, Smart factory systems, Informatik-Spektrum 38 (2015) 230–235.

[13] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. Hashem, A. Siddiqa, I. Yaqoob, Big iot data analytics: Architecture, opportunities, and open research challenges, IEEE Access (2017).

[14] M. H. Rehman, A. Batool, et al., Big data analytics in mobile and cloud computing environments, in: Innovative Research and Applications in Next-Generation High Performance Computing, IGI Global, 2016, pp. 349–367.

[15] K. Geng, L. Liu, Research of construction and application of cloud storage in the environment of industry 4.0, in: International Conference on Industrial IoT Technologies and Applications, Springer, pp. 104–113.

[16] M. M. Gaber, A. Aneiba, S. Basurra, O. Batty, A. M. Elmisery, Y. Kovalchuk, M. H. Rehman, Internet of things and data mining: From applications to techniques and systems, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery (2019) e1292.

[17] M. H. Rehman, L. C. Sun, T. Y. Wah, M. K. Khan, Towards next-generation heterogeneous mobile data stream mining applications: opportunities, challenges, and future research directions, Journal of Network and Computer Applications (2016).

[18] M. Ehret, J. Wirtz, Unlocking value from machines: business models and the industrial internet of things, Journal of Marketing Management 33 (2017) 111–130.

[19] L. Da Xu, W. He, S. Li, Internet of things in industries: A survey, IEEE Transactions on industrial informatics 10 (2014) 2233–2243.

[20] S. Mumtaz, A. Alsohaily, Z. Pang, A. Rayes, K. F. Tsang, J. Rodriguez, Massive internet of things for industrial applications: Addressing wireless iiot connectivity challenges and ecosystem fragmentation, IEEE Industrial Electronics Magazine 11 (2017) 28–33.

[21] F. Al-Turjman, S. Alturjman, Context-sensitive access in industrial internet of things (iiot) healthcare applications, IEEE Transactions on Industrial Informatics 14 (2018) 2736–2744.

[22] R. Chaudhary, G. S. Aujla, S. Garg, N. Kumar, J. J. Rodrigues, Sdn-enabled multi-attribute-based secure communication for smart grid in iiot environment, IEEE Transactions on Industrial Informatics 14 (2018) 2629–2640.

[23] C. Perera, C. H. L. Member, S. Jayawardena, M. Chen, Context-aware computing in the internet of things: a survey on internet of things from industrial market perspective, arXiv preprint arXiv:1502.00164 (2015).

[24] C. Perera, C. H. Liu, S. Jayawardena, The emerging internet of things marketplace from an industrial perspective: A survey, IEEE Transactions on Emerging Topics in Computing 3 (2015) 585–598.

[25] E. Ahmed, I. Yaqoob, I. A. T. Hashem, I. Khan, A. I. A. Ahmed, M. Imran, A. V. Vasilakos, The role of big data analytics in internet of things, Computer Networks 129 (2017) 459–471.

[26] C.-W. Tsai, C.-F. Lai, H.-C. Chao, A. V. Vasilakos, Big data analytics: a survey, Journal of Big Data 2 (2015) 21.

[27] Y. Chen, H. Chen, A. Gorkhali, Y. Lu, Y. Ma, L. Li, Big data analytics and big data science: a survey, Journal of Management Analytics 3 (2016) 1–42.

[28] M. K. Saggi, S. Jain, A survey towards an integration of big data analytics to big insights for value-creation, Information Processing and Management 54 (2018) 758 – 790. In (Big) Data we trust: Value creation in knowledge organizations.

[29] J. Wang, Y. Wu, N. Yen, S. Guo, Z. Cheng, Big data analytics for emergency communication networks: A survey, IEEE Communications Surveys & Tutorials 18 (2016) 1758–1778.

[30] H. Hu, Y. Wen, T.-S. Chua, X. Li, Toward scalable systems for big data analytics: A technology tutorial, IEEE access 2 (2014) 652–687.

[31] B. Li, M. C. Kisacikoglu, C. Liu, N. Singh, M. Erol-Kantarci, Big data analytics for electric vehicle integration in green smart cities, IEEE Communications Magazine 55 (2017) 19–25.

[32] M. H. Rehman, E. Ahmed, I. Yaqoob, I. A. T. Hashem, M. Imran, S. Ahmad, Big data analytics in industrial iot using a concentric computing model, IEEE Communications Magazine 56 (2018) 37–43.

[33] S. Radhakrishnan, S. Kamarthi, Convergence and divergence in academic and industrial interests on iot based manufacturing, in: Big Data (Big Data), 2016 IEEE International Conference on, IEEE, pp. 2051–2056.

[34] D. Georgakopoulos, P. P. Jayaraman, Internet of things: from internet scale sensing to smart services, Computing 98 (2016) 1041–1058.

[35] I. Lee, K. Lee, The internet of things (iot): Applications, investments, and challenges for enterprises, Business Horizons 58 (2015) 431–440.

[36] D. Georgakopoulos, P. P. Jayaraman, M. Fazia, M. Villari, R. Ranjan, Internet of things and edge cloud computing roadmap for manufacturing, IEEE Cloud Computing 3 (2016) 66–73.

[37] R. F. Babiceanu, R. Seker, Big data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook, Computers in Industry 81 (2016) 128–137.

[38] L. Wang, M. Törngren, M. Onori, Current status and advancement of cyber-physical systems in manufacturing, Journal of Manufacturing Systems 37 (2015) 517–527.

[39] H. S. Kang, J. Y. Lee, S. Choi, H. Kim, J. H. Park, J. Y. Son, B. H. Kim, S. Do Noh, Smart manufacturing: Past research, present findings, and future directions, International Journal of Precision Engineering and Manufacturing-Green Technology 3 (2016) 111–128.

[40] R. Addo-Tenkorang, P. T. Helo, Big data applications in operations/supply-chain management: A literature review, Computers & Industrial Engineering 101 (2016) 528–543.

[41] T.-M. Choi, H. K. Chan, X. Yue, Recent development in big data analytics for business operations and risk management, IEEE transactions on cybernetics 47 (2017) 81–92.

[42] C. Chang, S. N. Srirama, R. Buyya, Mobile cloud business process management system for the internet of things: a survey, ACM Computing Surveys (CSUR) 49 (2016) 70.

[43] K. Zhou, T. Liu, L. Liang, From cyber-physical systems to industry 4.0: make future manufacturing become possible, International Journal of Manufacturing Research 11 (2016) 167–188.

[44] J. Lee, H. D. Ardakani, S. Yang, B. Bagheri, Industrial big data analytics and cyber-physical systems for future maintenance & service innovation, Procedia CIRP 38 (2015) 3–7.

[45] F.-T. Cheng, H. Tieng, H.-C. Yang, M.-H. Hung, Y.-C. Lin, C.-F. Wei, Z.-Y. Shieh, Industry 4.1 for wheel machining automation, IEEE Robotics and Automation Letters 1 (2016) 332–339.

[46] L. Cheng, T. Wang, X. Hong, Z. Wang, J. Wang, J. Liu, A study on the architecture of manufacturing internet of things, International Journal of Modelling, Identification and Control 23 (2015) 8–23.

[47] R. Y. Zhong, C. Xu, C. Chen, G. Q. Huang, Big data analytics for physical internet-based intelligent manufacturing shop floors, International Journal of Production Research 55 (2017) 2610–2621.

[48] F. Tao, Y. Cheng, L. Da Xu, L. Zhang, B. H. Li, Cciot-cmfg: cloud computing and internet of things-based cloud manufacturing service system, IEEE Transactions on Industrial Informatics 10 (2014) 1435–1442.

[49] M. S. Hossain, G. Muhammad, Cloud-assisted industrial internet of things (iiot)–enabled framework for health monitoring, Computer Networks 101 (2016) 192–202.

[50] H. Gupta, A. V. Dastjerdi, S. K. Ghosh, R. Buyya, ifogsim: A toolkit for modeling and simulation of resource management techniques in internet of things, edge and fog computing environments, arXiv preprint arXiv:1606.02007 (2016).

[51] M. H. Rehman, P. P. Jayaraman, C. Perera, The emergence of edge-centric distributed iot analytics platforms, in:

Internet of Things, Chapman and Hall/CRC, 2017, pp. 213–228.

[52] K. Lin, F. Xia, G. Fortino, Data-driven clustering for multimedia communication in internet of vehicles, Future Generation Computer Systems 94 (2019) 610–619.

[53] P. Pace, G. Aloi, R. Gravina, G. Caliciuri, G. Fortino, A. Liotta, An edge-based architecture to support efficient applications for healthcare industry 4.0, IEEE Transactions on Industrial Informatics 15 (2019) 481–489.

[54] N. Hassan, S. Gillani, E. Ahmed, I. Yaqoob, M. Imran, The role of edge computing in internet of things, IEEE Communications Magazine (2018) 1–6.

[55] C. Ji, Q. Shao, J. Sun, S. Liu, L. Pan, L. Wu, C. Yang, Device data ingestion for industrial big data platforms with a case study, Sensors 16 (2016) 279.

[56] H. Wang, M. Li, Y. Bu, J. Li, H. Gao, J. Zhang, Cleanix: A parallel big data cleaning system, ACM SIGMOD Record 44 (2016) 35–40.

[57] A. F. Haryadi, J. Hulstijn, A. Wahyudi, H. van der Voort, M. Janssen, Antecedents of big data quality: An empirical examination in financial service organizations, in: Big Data (Big Data), 2016 IEEE International Conference on, IEEE, pp. 116–121.

[58] J. LeFevre, R. Liu, C. Inigo, L. Paz, E. Ma, M. Castellanos, M. Hsu, Building the enterprise fabric for big data with vertica and spark integration, in: Proceedings of the 2016 International Conference on Management of Data, ACM, pp. 63–75.

[59] Y. Wang, L. Kung, W. Y. C. Wang, C. G. Cegielski, Integrated big data analytics-enabled transformation model: Application to health care, Information & Management (2017).

[60] G. Press, Cleaning big data: Most time-consuming, least enjoyable data science task, survey says, Online, 2016.

[61] P. Wiener, M. Stein, D. Seebacher, J. Bruns, M. Frank, V. Simko, S. Zander, J. Nimis, Biggis: a continuous refinement approach to master heterogeneity and uncertainty in spatio-temporal big data (vision paper), in: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, p. 8.

[62] B. Ong, R. Wen, A. N. Zhang, Data blending in manufacturing and supply chains, in: Big Data (Big Data), 2016 IEEE International Conference on, IEEE, pp. 3773–3778.

[63] D. García-Gil, J. Luengo, S. García, F. Herrera, Enabling smart data: Noise filtering in big data classification, arXiv preprint arXiv:1704.01770 (2017).

[64] M. Bai, X. Wang, J. Xin, G. Wang, An efficient algorithm for distributed density-based outlier detection on big data, Neurocomputing 181 (2016) 19–28.

[65] L. Stojanovic, M. Dinic, N. Stojanovic, A. Stojadinovic, Big-data-driven anomaly detection in industry (4.0): An approach and a case study, in: Big Data (Big Data), 2016 IEEE International Conference on, IEEE, pp. 1647–1652.

[66] M. Niño, F. Sáenz, J. M. Blanco, A. Illarramendi, Requirements for a big data capturing and integration architecture in a distributed manufacturing scenario, in: Industrial Informatics (INDIN), 2016 IEEE 14th International Conference on, IEEE, pp. 1326–1329.

[67] R. Wang, D. Sun, G. Li, M. Atif, S. Nepal, Logprov: Logging events as provenance of big data analytics pipelines with trustworthiness, in: IEEE Conference on Big Data.

[68] K. Yamanishi, K. Miyaguchi, Detecting gradual changes from data stream using mdl-change statistics, in: Big

Data (Big Data), 2016 IEEE International Conference on, IEEE, pp. 156–163.

[69] X. Cheng, L. Fang, X. Hong, L. Yang, Exploiting mobile big data: Sources, features, and applications, IEEE Network 31 (2017) 72–79.

[70] A. A. F. Saldivar, C. Goh, W.-n. Chen, Y. Li, Self-organizing tool for smart design with predictive customer needs and wants to realize industry 4.0, in: Evolutionary Computation (CEC), 2016 IEEE Congress on, IEEE, pp. 5317–5324.

[71] A. A. F. Saldivar, C. Goh, Y. Li, Y. Chen, H. Yu, Identifying smart design attributes for industry 4.0 customization using a clustering genetic algorithm, in: Automation and Computing (ICAC), 2016 22nd International Conference on, IEEE, pp. 408–414.

[72] C. Yang, S. Lan, W. Shen, G. Q. Huang, X. Wang, T. Lin, Towards product customization and personalization in iot-enabled cloud manufacturing, Cluster Computing (2017) 1–14.

[73] D. Zurita, M. Delgado, J. A. Carino, J. A. Ortega, G. Clerc, Industrial time series modelling by means of the neo-fuzzy neuron, IEEE Access 4 (2016) 6151–6160.

[74] C. Gamarra, J. M. Guerrero, E. Montero, A knowledge discovery in databases approach for industrial microgrid planning, Renewable and Sustainable Energy Reviews 60 (2016) 615–630.

[75] H. Fleischmann, J. Kohl, J. Franke, A. Reidt, M. Duchon, H. Krcmar, Improving maintenance processes with distributed monitoring systems, in: Industrial Informatics (INDIN), 2016 IEEE 14th International Conference on, IEEE, pp. 377–382.

[76] D. L. Nuñez, M. Borsato, An ontology-based model for prognostics and health management of machines, Journal of Industrial Information Integration (2017).

[77] E. G. Nabati, K.-D. Thoben, Data driven decision making in planning the maintenance activities of off-shore wind energy, Procedia CIRP 59 (2017) 160–165.

[78] J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas, A. V. Vasilakos, A manufacturing big data solution for active preventive maintenance, IEEE Transactions on Industrial Informatics (2017).

[79] R. A. A. Habeeb, F. Nasaruddin, A. Gani, I. A. T. Hashem, E. Ahmed, M. Imran, Real-time big data processing for anomaly detection: A survey, International Journal of Information Management (2018).

[80] M. H. Rehman, V. Chang, A. Batool, T. Y. Wah, Big data reduction framework for value creation in sustainable enterprises, International Journal of Information Management 36 (2016) 917–928.

[81] J. R. B. Mozafari, S. W. S. Menon, N. K. H. B. S. Chakraborty, Y. M. R. M. K. Bachhav, Snappydata: Streaming, transactions, and interactive analytics in a unified engine (2016).

[82] P. P. De Silva, P. A. De Silva, Ipanera: An industry 4.0 based architecture for distributed soil-less food production systems, in: Manufacturing & Industrial Engineering Symposium (MIES), IEEE, pp. 1–5.

[83] V. K. Vavilapalli, A. C. Murthy, C. Douglas, S. Agarwal, M. Konar, R. Evans, T. Graves, J. Lowe, H. Shah, S. Seth, et al., Apache hadoop yarn: Yet another resource negotiator, in: Proceedings of the 4th annual Symposium on Cloud Computing, ACM, p. 5.

[84] J. Dean, S. Ghemawat, Mapreduce: simplified data processing on large clusters, Communications of the ACM 51 (2008) 107–113.

[85] X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen, et al., Mllib: Machine learning in apache spark, Journal of Machine Learning Research 17 (2016) 1–7.

[86] H. Lee, Framework and development of fault detection classification using iot device and cloud environment, Journal of Manufacturing Systems (2017).

[87] H. Lee, R. Grosse, R. Ranganath, A. Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in: Proceedings of the 26th annual international conference on machine learning, ACM, pp. 609–616.

[88] G. E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural computation 18 (2006) 1527–1554.

[89] Y. Zhang, S. Ren, Y. Liu, S. Si, A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products, Journal of Cleaner Production 142 (2017) 626–641.

[90] M. Takikawa, Innovation in railway maintenance utilizing information and communication technology (smart maintenance initiative), Japan Railway & Transport Review (2016).

[91] J. Cao, H. Cui, H. Shi, L. Jiao, Big data: A parallel particle swarm optimization-back-propagation neural network algorithm based on mapreduce, PLoS One 11 (2016) e0157551.

[92] M. Dopico, A. Gomez, D. De la Fuente, N. García, R. Rosillo, J. Puche, A vision of industry 4.0 from an artificial intelligence point of view, in: Proceedings on the International Conference on Artificial Intelligence (ICAI), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), p. 407.

[93] S. Wang, J. Wan, D. Zhang, D. Li, C. Zhang, Towards smart factory for industry 4.0: A self-organized multi-agent system with big data based feedback and coordination, Computer Networks 101 (2016) 158–168.

[94] T. Y. Berger-Wolf, Research: Computational ecology, social network analysis, data science., Journal of Parallel and Distributed Computing 93 (2016) 66–86.

[95] M. H. Rehman, C. S. Liew, A. Abbas, P. P. Jayaraman, T. Y. Wah, S. U. Khan, Big data reduction methods: A survey, Data Science and Engineering (2016) 1–20.

[96] T. Töniges, S. Ötting, B. Wrede, G. W. Maier, G. Sagerer, An emerging decision authority: Adaptive cyber-physical system design for fair human-machine interaction and decision processes, in: Cyber-physical systems: Foundations, principles, and applications, 2017.

[97] S. J. Oks, A. Fritzsche, K. M. Möslein, An application map for industrial cyber-physical systems, in: Industrial Internet of Things, Springer, 2017, pp. 21–46.

[98] M. Martellini, S. Abaimov, S. Gaycken, C. Wilson, Assessing cyberattacks against wireless networks of the next global internet of things revolution: Industry 4.0, in: Information Security of Highly Critical Wireless Networks, Springer, 2017, pp. 63–69.

[99] K. Jensen, H. T. Nguyen, T. Van Do, A. Årnes, A big data analytics approach to combat telecommunication vulnerabilities, Cluster Computing (2017) 1–12.

[100] H. Igor, J. Bohuslava, J. Martin, Proposal of communication standardization of industrial networks in industry 4.0, in: Intelligent Engineering Systems (INES), 2016 IEEE 20th Jubilee International Conference on, IEEE, pp. 119–124.

[101] A. Dagnino, D. Cox, Industrial analytics to discover knowledge from instrumented networked machines., in: SEKE, pp. 86–89.

[102] C. M. de Farias, L. Pirmez, G. Fortino, A. Guerrieri, A multi-sensor data fusion technique using data correlations

among multiple applications, Future Generation Computer Systems 92 (2019) 109–118.

[103] P.-Y. Wu, C.-W. Cheng, C. D. Kaddi, J. Venugopalan, R. Hoffman, M. D. Wang, -omic and electronic health record big data analytics for precision medicine., IEEE transactions on bio-medical engineering 64 (2017) 263–273.

[104] J. Wan, S. Tang, D. Li, M. Imran, C. Zhang, C. Liu, Z. Pang, Reconfigurable smart factory for drug packing in healthcare industry 4.0, IEEE transactions on industrial informatics 15 (2019) 507–516.

[105] I. Yaqoob, I. A. T. Hashem, A. Ahmed, S. A. Kazmi, C. S. Hong, Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges, Future Generation Computer Systems 92 (2019) 265 – 275.

[106] I. Yaqoob, E. Ahmed, M. H. ur Rehman, A. I. A. Ahmed, M. A. Al-garadi, M. Imran, M. Guizani, The rise of ransomware and emerging security challenges in the internet of things, Computer Networks 129 (2017) 444–458.

[107] Y.-W. Ma, Y.-C. Chen, J.-L. Chen, Sdn-enabled network virtualization for industry 4.0 based on iots and cloud computing, in: Advanced Communication Technology (ICACT), 2017 19th International Conference on, IEEE, pp. 199–202.

[108] M. Gupta, N. Singla, Evolution of cloud in big data with hadoop on docker platform, in: Privacy and Security Policies in Big Data, IGI Global, 2017, pp. 41–62.

[109] T.-H. Dang-Ha, D. Roverso, R. Olsson, Graph of virtual actors (gova): A big data analytics architecture for iot, in: Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on, IEEE, pp. 162–169.

[110] M. Fokaefs, C. Barna, R. Veleda, M. Litoiu, J. Wigglesworth, R. Mateescu, Enabling devops for containerized data-intensive applications: an exploratory study, in: Proceedings of the 26th Annual International Conference on Computer Science and Software Engineering, IBM Corp., pp. 138–148.

[111] E. Wagner, Microservices as a Manufacturing Execution System Architecture, Ph.D. thesis, Ulm University, 2016.

[112] S. A. Jacobs, A. Dagnino, Large-scale industrial alarm reduction and critical events mining using graph analytics on spark, in: Big Data Computing Service and Applications (BigDataService), 2016 IEEE Second International Conference on, IEEE, pp. 66–71.

[113] K. Salah, M. H. U. Rehman, N. Nizamuddin, A. Al-Fuqaha, Blockchain for ai: Review and open research challenges, IEEE Access 7 (2019) 10127–10149.

[114] N. Hassan, S. Gillani, E. Ahmed, I. Yaqoob, M. Imran, The role of edge computing in internet of things, IEEE Communications Magazine (2018) 1–6.

[115] S. El Kafhali, K. Salah, Efficient and dynamic scaling of fog nodes for iot devices, The Journal of Supercomputing 73 (2017) 5261–5284.

[116] R. Almadhoun, M. Kadadha, M. Alhemeiri, M. Alshehhi, K. Salah, A user authentication scheme of iot devices using blockchain-enabled fog nodes, in: Proceedings of the 15th ACS/IEEE International Conference on Computer Systems and Applications (ACS/IEEE AICCSA 2018), Aqaba, Jordan, October 28, 2018.