American Journal of Information Science and Computer Engineering
Vol. 2, No. 6, 2016, pp. 70-78
http://www.aiscience.org/journal/ajisce
ISSN: 2381-7488 (Print); ISSN: 2381-7496 (Online)

Big Data Analytics and Cloud Computing in Internet of Things

Lidong Wang[1, *], Cheryl Ann Alexander[2]

[1]Department of Engineering Technology, Mississippi Valley State University, Itta Bena, Mississippi, USA
[2]Technology and Healthcare Solutions, Inc., Itta Bena, Mississippi, USA

Abstract

Internet of Things (IoT) can be sensors, radio frequency identification (RFID) devices, or smart objects with the Internet connectivity over physical IP for transmitting data to the network. IoT generates big data with noise, variety, heterogeneity, high redundancy, and unstructured features. There are a lot of challenges in processing IoT. Big Data analytics and cloud computing are powerful tools for analyzing complicated data generated from IoT. This paper introduces general IoT, RFID, Big Data analytics (BDA); presents the progress of Big Data analytics for IoT and IoT data processing based on cloud computing. Challenges in these areas are also discussed.

Keywords

Big Data Analytics, Internet of Things (IoT), RFID, Cloud Computing, Machine to Machine (M2M), Wireless Sensor Networks, Machine Learning, Data Mining, Networking and Communications

Received: August 11, 2016 / Accepted: August 22, 2016 / Published online: November 2, 2016

1. Introduction

There are different definitions of the Internet of Things (IoT), but it is generally accepted as defined by the RFID group as "the worldwide network of interconnected objects uniquely addressable based on standard communications protocols" [1]. IoT was initially referred to uniquely identifiable, interoperable, and connected objects with radio frequency identification (RFID). Later on, researchers related IoT with more technologies such as sensors, actuators, GPS devices, and mobile devices. The integration of sensors/actuators, RFID tags, and communication technologies serves as the foundation of IoT [2].

For IoT, sensors and actuators embedded in physical objects – from roadways to pacemakers – are linked through wired and wireless networks, often using the same Internet Protocol (IP) that connects the Internet. 'Things' can be sensors, databases, and other devices or software. Sensors could include pacemakers, location identifiers such as global positioning system (GPS), and individual identification devices such as RFID tags [3]. In the IoT paradigm, many networking sensors are embedded into various devices and machines. Such sensors used in different fields may collect different kinds of data, such as environmental data, geographical data, astronomical data, and logistic data. Mobile equipment, transportation facilities, public facilities, and home appliances can be data acquisition equipment in IoT [4].

Early IoT applications are based on RFID and wireless sensor network (WSN) technologies and deliver tangible benefits in several areas including manufacturing, logistics, trade, retail, and green/sustainable applications, etc. [5]. In the field of the IoT, more than 30 million networked sensor nodes were functioning in the transportation, automotive, industrial, utilities, and retail sectors. The number of these sensors was increasing at a rate of more than 30 percent per

* Corresponding author
E-mail address: lwang22@students.tntech.edu (Lidong Wang)

year [6]. Despite having played a significant role in the Industry 4.0 era, IoT is faced with the challenge of how to ingest large-scale heterogeneous and multi-type device data. Smart manufacturing has become a vital component of manufacturing in the Industry 4.0 era. With the development of smart manufacturing technology, it can be foreseen that IoT will increase the scale of data to an unprecedented level. More effective approaches for resolving record storage and queries in a big data environment are required [7].

IoT can generate big data. RFID tags generate volumes and volumes of data. As a result, digital processing becomes a requirement of feasibility. The velocity of data associated with the 'Internet of Things' explodes as sensors can continuously capture data. The variety of data associated with the 'Internet of Things' also is expansive as the types of sensors and the different sources of data expand. The veracity of data in the 'Internet of Things' may also be improving as the quality of sensor and other data improves over time [3]. In addition to very large data sets, big data can also be a mix of structured data (relational database), unstructured data (human language text), semi-structured data (RFID or XML), and streaming data (from machines, sensors, Web applications, and social media). Structured data from applications is a common form of big data although it's not new. 10 to 99 terabytes are the big data norm today [8]. IoT can comprise billions of devices that can sense, communicate, compute, and potentially actuate. Data streams coming from these devices challenge the traditional approaches to data management and contribute to the paradigm of big data. If Wal-Mart operates RFID on an item level, it is expected to generate about seven terabytes (TB) of data every day [9].

In an IoT big data system, bulk amounts of data are organized in the form of NoSQL databases. The IoT big data is a spatiotemporal database that depends on the time and location; more numbers of rows are there along with less number of columns. So the column oriented data-depository can greatly improve the performance of IoT big data in data accessing and query processing. The heterogeneous IoT big data cannot be stored in any relation database. Therefore, IoT big data cell (NoSQL database) may be used to resolve the storage limitation and constraints of relational database [10].

The organization of the paper is as follows: the next section introduces Internet of Things (general Internet of Things and RFID); Section 3 introduces Big Data analytics; Section 4 introduces Big Data analytics and cloud computing for Internet of Things; and the final section is conclusion.

2. Internet of Things

2.1. General Internet of Things

IoT forms a communicating-actuating network of a large amount of things including RFID tags, mobile phones, sensors, and actuators, etc. [11]. The data generated from IoT has the following features [4]:

- Large-scale data: Masses of data acquisition equipment are distributed. For analysis and processing, not only the currently acquired data, but also the historical data within a certain time frame should be stored. Therefore, IoT generates large-scale data.
- Heterogeneity: Because of the variety of data acquisition devices, the acquired data is also different, which results in data heterogeneity.
- Strong time and space correlation: Every data acquisition device is placed at a specific geographic location and every piece of data has a time stamp. The time and space correlation is an important property of data from IoT.
- Effective data accounts for only a small portion of big data: a great quantity of noises may occur during the acquisition and transmission of data in IoT. In some situations, only a small amount of abnormal data is valuable. For example, a small amount of traffic video frames that capture the violation of traffic regulations and traffic accidents are really valuable.

A foundational technology for IoT is RFID. People can identify, track, and monitor any objects attached with RFID tags automatically. Another foundational technology for IoT is the wireless sensor networks (WSNs) [2]. Some technologies associated with IoT are summarized as follows [2]:

- Identification and Tracking Technologies: RFID systems, barcode, and intelligent sensors.
- Communication Technologies.
- Networks: IoT involves a number of heterogeneous networks such as WSNs. These networks must be revised before they can be applied to IoT. The reason is that things in IoT often have diverse communication and computation capabilities. In contrast, nodes in WSNs typically have similar requirements for hardware and network communication. In addition, the IoT network uses the Internet to support information exchange and data communication. In contrast, WSNs do not have to involve the Internet for communication.
- Cloud computing
- Service Management in IoT.

IoT has three unique features: intermittent sensing, regular data collection, and sense-compute-actuate (SCA) loops [9]. Telematics is a prime example of an industry harnessing the power of mobile connectivity and IoT. Complex data types typically in IoT applications can be modeled and represented more efficiently using JSON (JavaScript Object Notation) documents, rather than tables [12]. IoT can be divided into five layers as shown in Table 1 [13].

Table 1. IoT five-layer architecture.

Layers	Description
Perception Layer (device layer)	Composed of physical objects and sensor devices
Network Layer (transport layer)	The medium of data transition may be wired or wireless, and the using technology can be Wi-Fi, 3G, Zig-Bee, and Z-Wave, etc.
Middleware Layer	Responsible for storing, analyzing, and processing the information of objects that received from the network layer and linked to the database.
Application Layer	Offers inclusive management of application that relies on the objects information processed in the middleware layer.
Business Layer	Business layer likes a manager of IoT. The management includes applications, relevant system model, and services.

WSNs are being enabled by the increasing availability of sensors and advances in wireless technologies, hardware, and the use of IP for connecting resource constrained devices. The use of micro IP stacks has enabled constrained devices to be connected to IoT. IoT is characterized by an interconnected set of individually addressed and constrained (possibly autonomous) devices in a distributed system, with sensing/active devices for physical phenomena, data collection, and applications. However, the potential of WSNs is limited by the relatively low number deployed and the difficulties imposed by their heterogeneous nature and limited (or proprietary) development environments and interfaces. The constrained nature of WSN nodes in processing power, memory and energy consumption is a challenge. A set of requirements were proposed for achieving a pervasive, integrated information system of WSNs and associated services. An architecture was presented, which considered the data flow from sensors through to services and provides a set of abstractions for the different types of sensors and services [14].

Industrial IoT (IIoT), a sub-paradigm of IoT, focuses more in safety-critical industrial applications. Compared with domestic IoT, IIoT has long product cycles, often operating in extreme conditions. IIoT generally needs to be integrated with other industrial systems from different vendors; while domestic IoT normally is a vertically integrated, single-vendor solution. IIoT must prevent unauthorized access while domestic IoT is more concerned with users' privacy issues. IIoT must be fault-tolerant and cannot assume continuous access to Internet or cloud; therefore, IIoT has to be autonomous and be able to function during network interruptions [11].

Interoperability is a key challenge in IoT. This is due to the intrinsic fabric of IoT: (1) high–dimensional, with the co-existence of many systems (devices, sensors, and equipment, etc.) in the environment that need to communicate and exchange information; (2) highly heterogeneous, where these vast systems are conceived by a lot of manufacturers and are designed for many different purposes and diverse application domains, making it extremely difficult to reach out for global agreements and widely accepted specification; (3) dynamic and non-linear; and (4) hard to describe/model due to existence of many data formats. Sustainable interoperability is needed in IoT. The framework for sustainable interoperability in IoT needs to address the following aspects: (1) management of interoperability in IoT; (2) dynamic interoperability technologies for IoT; (3) measurement of interoperability in IoT (need to quantify and/or qualify interoperability); (4) interaction and integration of IoT in the global Internet: IPv6 integration, global interoperability, and IoT-Cloud integration, etc. In other words, it is needed to address how to bridge billion of smart things globally, while respecting their specific constraints [15].

IoT, mobile computing (MC), pervasive computing (PC), wireless sensor networks (WSNs), and most recently, cyber physical systems (CPS) are five research communities. However, as technology and solutions progress in each of these fields there is an increasing overlap and merger of principles and research questions. Research in IoT, PC, MC, WSN and CPS often relies on underlying technologies such as real-time computing, machine learning, security, privacy, signal processing, and big data, etc. [16].

The spectrum of research required to achieve IoT at the scale requires significant research along many directions, which are highlighted in eight topic areas: massive scaling, architecture and dependencies, creating knowledge and big data, robustness, openness, security, privacy, and human-in-the-loop. As for human-in-the-loop, IoT applications will become more sophisticated when they proliferate. Many of these new applications will intimately involve humans, i.e., humans and things will operate synergistically. Human in-the-loop systems offer great opportunities to a broad range of applications, including energy management, health care, and automobile systems. However, modeling human behaviors is a challenge due to complicated physiological, psychological, and behavioral aspect of human beings. New research is needed to raise human-in-the-loop control to a central principle in system design. One vision of the future is that IoT becomes a utility with increased sophistication in sensing, actuation, communications, control, and in creating knowledge from vast amounts of data [16].

There are several techniques or tools for solving IoT data management challenges. They are: Big data, cloud computing, semantic sensor web, data fusion techniques, and middleware [13]. In addition, the following research trends [2] also should be our concerns:

- Developing green IoT technologies: There is a need to develop energy-efficient techniques that can reduce the consumed power by sensors.
- Developing context-aware IoT middleware solutions to better understand sensor data and help decide what data needs to be processed.
- Employing artificial intelligence techniques to create intelligent things or smart objects: Future IoT systems should have characteristics including “self-configuration, self-optimization, self-protection, and self-healing”.
- Combining IoT and cloud computing, implementing new models or platforms that provide “sensing as a service” on the cloud.

2.2. Radio Frequency Identification

With the IoT technologies such as RFID implemented in manufacturing sites, enormous data will be generated. Such data are so complex, abstract, and variable that it is difficult to make full use of the data that carry useful information [17]. RFID-enabled item-level tagging, is expected to generate not only huge operational and strategic data across the value chain of all industries, but also an impressive volume of RFID data [18].

EPCIS is an RFID event repository, which is one of the core component of the EPCglobal Architecture Framework. It helps store RFID event information and share the information among supply chain partners. Electronic Product Code (EPC) refers to a coding scheme for unambiguous code for the designation of physical goods. RFID and sensor technologies are the core technologies of future IoT [19]. RFID creates new possibilities and processes such as real time inventory and item-level process validation. As a result, a much larger volume of data is generated, and it is item-level data rather than transactional data. When a major retailer implements RFID on a significant portion of their products, the result can be literally billions of additional data points. Most existing enterprise systems were designed to handle transaction-level data (like a P.O. or shipment) and not designed to handle item-level data. What is needed is a new breed of application designed specifically to handle item-level RFID data, while integrating to existing legacy systems and processes. The new systems should be able to [20]:

- Translate item-level data into transaction-level data that existing systems can absorb.
- Filter and consolidate item-level data into meaningful business events, making efficient use of network bandwidth.
- Provide management-by-exception via rules-based monitoring of these enormous new data flows.
- Store item-level data in an EPICS-compliant database, organized in a hierarchical manner (i.e. consolidating local event data up to higher levels business event data as needed).
- Provide business intelligence and analytic tools designed specifically to leverage this granular item-level data.

The RFID-enabled logistics big data usually include some “noise” such as incomplete, redundant, and inaccurate records. The major noises in RFID-enabled logistics data come from redundant records. Thus, it is important to detect and remove the redundancy. However, current methods are not suitable for removing the above noises. The knowledge hidden in the RFID-enabled big data is sporadic. That means hundreds of RFID records may create a piece of information which indicates the detailed logic operations [21]. Therefore, raw RFID data are typically of low quality and may contain many anomalies because of physical device limitations and different types of environmental noise. RFID data poses many challenges for data analysis: (1) RFID data are inherently noisy and redundant; (2) RFID data are temporal, streaming, high volume, and must be processed on the fly [6]. Several key procedures are proposed: an RFID-Cuboid cleansing algorithm was presented for detecting and removing the noise data from the logistics dataset; an RFID-Cuboid compression algorithm was demonstrated for reducing the storage space and enhancing information granularity; and an RFID-Cuboid classification algorithm was reported for clustering the cuboids according to the practical applications/considerations [21].

3. Big Data Analytics

Big data comes from a variety of sources, in very large amounts, and often in real-time settings. This trend is largely driven by the pervasive diffusion and adoption of mobile devices, social media tools, and IoT enabled by RFID and other RF-related tracking and sensor devices [1]. Table 2 [22] lists the classification of big data.

Table 2. Big data classification.

Classification	Description
Data Sources	• Web & Social: Generating data via URL to share or exchange information in virtual communities and networks, such as blogs, Facebook, and Twitter. • Machine: automatically generating data from computers, medical devices, or other machines. • Sensing: generating data from sensing devices • Transactions: Transaction data, such as financial and work data. • IoT: Internet of things produces huge amounts of data.
Content Formats	• Structured: Structured data are often managed SQL, a programming language created for managing and querying data in RDBMS. • Semi-structured: not following a conventional database system; in the form of structured data that are not organized in relational database models. • Unstructured: such as text messages, videos, and social media data; not following a specified format.
Data Stores	• Document-oriented: A document-oriented data store is similar to a record or row in a relational database but is more flexible and can retrieve documents based on their contents. • Column-oriented: A column-oriented database stores its content in columns aside from rows, with attribute values belonging to the same column stored contiguously. • Graph based: storing and representing data that utilize a graph model with nodes, edges, and properties related to one another through relations. • Key-value: Key-value is an alternative relational database system that stores and accesses data designed to scale to a very large size.
Data Staging	• Cleaning: identifying incomplete and unreasonable data. • Normalization: structuring database schema to minimize redundancy. • Transform: transforming data into a form suitable for analysis.
Data Processing	• Batch: MapReduce-based systems have been adopted for long-running batch jobs. • Real time: such as simple scalable streaming system (S4).

Although Hadoop has become a mainstay of big data analytics platforms, it remains far from mature. First, Hadoop must integrate with real-time massive data collection & transmission and provide faster processing beyond the batch-processing paradigm. Second, Hadoop provides a concise user programming interface, while hiding the complex background execution. In some senses, this simplicity causes poor performance. It is difficult for current and mature batch-processing paradigms to adapt to the rapidly growing data volume and the substantial real-time requirements. In-situ analysis avoids the overhead of file transfer to the centralized storage infrastructure to improve real-time performance. Due to the value-sparse feature of big data, a new data analysis mechanism should adopt dimensionality reduction or sampling-based data analysis to reduce the amount of data to be analyzed [6]. MapReduce is a programming and processing model in big data. Some MapReduce projects and related software are shown in Table 3.

Table 3. Some MapReduce projects and related software.

Software	Brief Description
Hive	Offers a warehouse structure in HDFS
Pig	Involves a high-level scripting language (Pig Latin) and offers a run-time platform allowing users to execute MapReduce on Hadoop
Hbase	Scalable distributed database supporting structured data storage for large tables
Spark™	A fast computation engine for Hadoop data
YARN	A new Apache–Hadoop–MapReduce framework
Cassandra	A scalable multi-master database with no single point of failure
Zookeeper™	High-performance service to coordinate the processes of distributed applications; a distributed service with master and slave nodes and stores configuration information
Madout™	A machine-learning and data-mining library that can be executed in a distributed mode and is executable by MapReduce

Approaches to handling big data workloads are such as: (1) classical data warehouses; (2) batch processing (e.g., Apache Hadoop); (3) real-time processing (e.g., Twitter Storm); and (4) edge computing. Real-time Processing Systems such as Apache Storm are much better suited to processing large amounts of streaming data than Hadoop-based systems. The advantage of these systems is that they can process much more data per second. However, Storm does not provide high-level query languages for data analysis such as Pig or Hive. Edge computing covers a wide range of technologies such as distributed data storage, mobile data acquisition, and fog computing that extends the cloud computing paradigm to the edge of the network. The edge servers are globally distributed to ensure consistent low latency when users request a page. There are different big data challenges such as real-time (velocity), scale-out (volume), and edge processing (variety and low latency combined with connectivity challenges). Many actually require data ingestion from multiple sources (variety) and at the same time—especially within the IoT domain—involve some real-time aspect or a direct feedback component [4].

Joining big data with traditional data is another path to value. For example, so-called 360-degree views of customers and other business entities are more complete and bigger when based on both traditional enterprise data and big data. Streaming big data is easy to capture, but tough to process in real time. Interest is high in distributed file systems and distributed analytic processing. In-database analytics takes processing to the data, instead of vice versa. Reducing disk I/O increases the performance of data intense processes. As for in-memory databases, one way to get high performance (in the sense of fast data access) from a database is to manage it in server memory, thereby eliminating disk I/O and other bottlenecks. A cloud can be central to a data management strategy [8].

A heterogeneous device data ingestion model was proposed for an Industrial Big Data Platform (IBDP). The model includes device templates and four strategies for data synchronization, data slicing, data splitting and data indexing, respectively. Device data from multiple sources can be ingested using this heterogeneous device data ingestion model, which is verified on the IBDP [7].

The existing relational database technologies are inadequate to handle IoT-generated data due to the limited processing speed and the significant storage expansion cost. Thus, big data processing technologies, which are normally based on distributed file systems, distributed database management, and parallel processing technologies, have arisen as a core technology to implement IoT-generated data repositories. A sensor-integrated RFID data repository-implementation model was proposed using MongoDB, a popular big data-savvy document-oriented database system and a document-oriented NoSQL database that offers high performance and scalability. A MongoDB-based RFID/sensor data repository was designed. It can integrate and store RFID and sensor data by referencing event types of Electronic Product Code Information Services (EPCIS) in the EPCglobal network [19].

4. Big Data Analytics and Cloud Computing for Internet of Things

4.1. Cloud Computing for IoT

IoT and mobility provide sensors that can sense in real-time even while moving. These sensors will produce big data that is high volume, high variety, and high velocity. An elastic infrastructure such as the cloud needs to be used to process the big data. In other terms, the cloud binds to the IoT. Cloud computing consists of three main layers or model: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). In addition, some other layers are such as Sensing as a service (SaaS), Data as a Service (DaaS), and Network as a Service (NaaS), etc. In general, all these models are called XaaS. The objective of IoT is to provide utility-based services such as Sensing as a service, Location-as-a-Service, and Traceability-as-a-Service. Very large scale sensor networks domain use cloud computing to process data in the cloud. Such data can be characterized as polymorphous, heterogeneous, large in quantity and time-limited. In a very large scale sensor network, managing the sensing resources and computational resources, and storing and processing these data are key challenges [9].

4.2. Big Data Analytics for IoT

IoT data can be big data. There are large volumes of data to read and write. The amount of data can be TB(terabytes), PB (petabytes), and even ZB(zettabyte). There are Heterogeneous data sources and data types to integrate. Data sources are diverse; for example, it is needed to integrate sensors data, cameras data, social media data, and so on. All these data are different in format: byte, binary, string, and number, etc. There is complex knowledge to extract. The knowledge is deeply hidden in large volumes of data. There are lots of challenges in processing IoT big data; the quantity of data is big but the quality is low due to different data sources and different types and representation forms (structured, semi-structured, and unstructured) [23].

Context-based devices have been a concern of the 'Internet of Things'. The 'Internet of Things' provides potential for capturing and generating context about relevant 'things'. If the 'things' are RFID tags representing inventory, then people would have some understanding and expectations regarding behavior because the tags are representing inventory. 'Big Data' also should be able to provide 'Big Context'. Sensor information operates in real time, speeding the velocity, resulting in a continuous monitoring of the 'Internet of Things' that is a source of 'Big Data' [3]. Semantic challenge is to extract the meaning of the information from massive volumes of unstructured dirty data [9].

Real-time data classification and clustering for agile data streams are challenging issues especially in IoT big data environment. A data classification process is expected to formulate fused data into multiple data groups. The fused data can be classified into multiple groups of having multiple event types such as machine status data, functional data, inventory data, production data, and product quality data, etc. Some nonlinear data classification process over the multilayer perceptron may be fit to the problem of large scale IoT big data classification. IoT big data management systems can be distributed computing systems that especially deal with semi-structured and unstructured IoT big data. Table 4 and Table 5 show an IoT big data management subsystem and an IoT big data layering architecture, respectively [10].

Table 4. IoT Big Data Management Subsystem.

Layers	Management Subsystems
Layer 1	IoT objects management (physical devices)
Layer 2	IoT big-data management
Layer 3	IoT intelligence management
Layer 4	IoT applications management

Table 5. Overall IoT Big Data Layering Architecture.

Layering Architecture	Representation
Application layer	IoT applications
Knowledge processing layer	IoT tools
Data management layer	IoT middleware
Transport layer	IoT network
Physical sensing layer	IoT objects

Machine to Machine (M2M) is regarded as the predecessor of the IoT in the industry. However, the evolution from M2M to the IoT is not only an issue of adding more devices. IoT will drive Big Data by providing more information, from many different sources, in real-time. This allows us to gain completely new perspectives on the environments around us. Big Data has five key capabilities for data management in IoT [12]:

- Creating rich and functional applications: Data management must support the development of functionally rich applications with complex data and algorithms.
- Unlocking business agility: The ability to support many new and frequently changing business requirements.
- Enabling a single point of truth & business convergence: Aggregate multiple views of related data from multiple systems into one consistent version of the data.
- Real-time operational insight: Support both operational as well as analytical applications from the same data source
- Enterprise-grade platform: Provide highly scalable, cloud-based, robust, and secure applications

The application of big analog data is the precursor to the rise of the Industrial Internet of Things (IIoT). By making machines smarter through local processing and communication, the IIoT will solve problems in ways that were previously inconceivable [24]. The result of rapid development of IoT/IIoT is that the enormous amount of collected data from different sources will have to be processed, analyzed, and visualized in a timely manner. This is where big data analytics (BDA) will fit in. In fact, BDA and IoT complement each other and develop as a double "helix". BDA on sensor-enabled operation data can improve energy efficiency and environmental performance, safety verification and assessment, and the monitoring of accidents and environment risks. In general, BDA requires heavy computational power. As people have observed in the HPC community, super computers have already been built with a hybrid CPU and GPU architecture to make use of the large pool of processing units in GPUs [11].

A holistic Big Data approach was proposed to excavate the frequent trajectory from massive RFID-enabled manufacturing data for supporting production logistics decision-makings. This approach comprises several key steps: warehousing for raw RFID data, cleansing mechanism for RFID big data, mining frequent patterns, as well as pattern interpretation and visualization [21].

For IoT applications, the obtained massive sensing data can be in various features, which is a challenge. Big Data analytics has been massive heterogeneous data analytics in nonlinear, high-dimensional, distributed, and parallel data processing. In Big Data techniques for IoT, an algorithm was proposed for anomaly detection in big sensor data. In particular, an algorithm of contextual anomaly detection was introduced to progress a point anomaly detection algorithm. A post-processing context-aware anomaly detection algorithm was proposed based on a multivariate clustering algorithm. A MapReduce methodology was also proposed to outline the sensor profiles used in the context detector [13].

4.3. Convergence Among Cloud Computing, Big Data Analytics and IoT

In the age of big data, hardware is evidently no longer the limiting factor in acquisition applications, but the management of acquired data is. It is simple to use cloud storage and cloud computing resources to create a single aggregation point for data coming in from a large number of embedded devices and provide access to that data from any group within an organization [24]. The convergence of IoT, Big Data and cloud lies in [25]:

- For Big Data, data collection is one of the main concern, and IoT can play important roles for data collection and data sharing
- Cloud offers Everything as a Service business model for IoT and Big Data.

Cloud services and Big Data approaches can be used to store and analyze IoT data to improve scalability and availability, which will be required for the billions of devices envisaged in IoT. It is necessary to enable WSNs to become extensions of the Internet infrastructure and take full advantage of cloud and Big Data services. The availability of increased storage and processing power at a lower cost with greater bandwidth has enabled a range of cloud computing services. In terms of IoT, this allows more sources of data to be collected and for the data to be held for a longer time and to be processed by powerful cloud based applications and Big Data techniques, e.g. HBase and MapReduce [14].

Big data storage and processing are considered as one of the main applications for cloud computing systems. Furthermore, the development of the IoT paradigm has advanced the research on M2M communications and enabled novel telemonitoring architectures for e-Health applications. However, there is a need for converging current decentralized cloud systems, general software for processing big data and IoT systems. Most IoT applications are based on M2M communication protocols between large numbers of heterogeneous and geographically distributed sensors. As a result, they need to handle hundreds (sometimes thousands) of sensor streams, and could directly benefit from the immense distributed storage capacities of cloud computing infrastructures. Furthermore, cloud infrastructures could boost the computational capacities of IoT applications. Also, several IoT services (e.g., large scale sensing experiments, smart city applications) could benefit from a utility-based delivery paradigm, which emphasizes the on-demand establishment and delivery of IoT applications over a cloud-based infrastructure. An M2M system was proposed based on a decentralized cloud architecture, general systems and remote telemetry units (RTUs) for e-Health applications. The system was built for big data processing of sensors information in the way that data can be aggregated to generate “virtual” sensors [5].

5. Conclusion

IoT and its tools cover RFID, wired/wireless sensors, networks, embedded systems, and computing and analytics such as cloud computing. IoT generates big data because of masses of data in a real timescale, often semi-structured or unstructured data, and valuable data only after being analyzed. Big data generated by IoT has some different features compared with general big data because of the different types of data collected.

RFID can identify, track, trace, and monitor objects. WSNs enable applications and services that may be located across the Internet from the sensing network. Both RFID and WSNs are foundational technologies for IoT. IoT has some challenges such as interoperability. Big data and cloud computing are powerful approaches to solving IoT challenges.

The following aspects can be future research topics: Engineered Resilient Systems (ERS) for IoT, artificial intelligence in IoT systems, green IoT technologies, context-aware IoT middleware for better understanding sensor data, real-time processing for IoT big data, and convergence of IoT, cloud computing and Big Data analytics.

References

[1] Riggins, F. J., & Wamba, S. F. (2015, January). Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on* (pp. 1531-1540). IEEE.

[2] Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, *10*(4), 2233-2243.

[3] O'Leary, D. E. (2013). BIG DATA', THE 'INTERNET OF THINGS'AND THE 'INTERNET OF SIGNS. *Intelligent Systems in Accounting, Finance and Management*, *20*(1), 53-65.

[4] Chen, M., Mao, S., & Liu, Y. (2014). Big data: a survey. *Mobile Networks and Applications*, *19*(2), 171-209.

[5] Suciu, G., Suciu, V., Martian, A., Craciunescu, R., Vulpe, A., Marcu, I., ... & Fratu, O. (2015). Big Data, Internet of Things and Cloud convergence–an architecture for secure e-health applications. *Journal of medical systems*, *39*(11), 1-8.

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

[7] Ji, C., Shao, Q., Sun, J., Liu, S., Pan, L., Wu, L., & Yang, C. (2016). Device Data Ingestion for Industrial Big Data Platforms with a Case Study. *Sensors*, *16*(3), 279.

[8] Russom, P. (2013). Managing big data. *TDWI Best Practices Report, TDWI Research*, 1-40.

[9] Zaslavsky, A., Perera, C., & Georgakopoulos, D. (2013). Sensing as a service and big data. *arXiv preprint arXiv:1301.0159*.

[10] Mishra, N., Lin, C. C., & Chang, H. T. (2015). A cognitive adopted framework for IoT big-data management and knowledge discovery prospective. *International Journal of Distributed Sensor Networks*, *2015*, 6.

[11] Wang, H., Osen, O. L., Li, G., Li, W., Dai, H. N., & Zeng, W. (2015, November). Big data and industrial internet of things for the maritime industry in northwestern Norway. In *TENCON 2015-2015 IEEE Region 10 Conference* (pp. 1-5). IEEE.

[12] Bosch and MongoDB, IoT and Big Data, A Joint Whitepaper by Bosch Software Innovations and MongoDB, August 2015.

[13] Aly, H., Elmogy, M., & Barakat, S. Big Data on Internet of Things: Applications, Architecture, Technologies, Techniques, and Future Directions.

[14] Tracey, D., & Sreenan, C. (2013, May). A holistic architecture for the internet of things, sensing services and big data. In *Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium on* (pp. 546-553). IEEE.

[15] Friess, P. (2013). *Internet of things: converging technologies for smart environments and integrated ecosystems*. River Publishers.

[16] Stankovic, J. A. (2014). Research directions for the internet of things. *IEEE Internet of Things Journal*, *1*(1), 3-9.

[17] Zhong, R. Y., Lan, S., Xu, C., Dai, Q., & Huang, G. Q. (2016). Visualization of RFID-enabled shopfloor logistics Big Data in Cloud Manufacturing. *The International Journal of Advanced Manufacturing Technology*, *84*(1-4), 5-16.

[18] Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234-246.

[19] Kang, Y. S., Park, I. H., Rhee, J., & Lee, Y. H. (2016). Mongodb-based repository design for iot-generated rfid/sensor big data. *IEEE Sensors Journal*, *16*(2), 485-497.

[20] Software Snapshot, FRPT Research, 2014, 11-13

[21] Zhong, R. Y., Huang, G. Q., Lan, S., Dai, Q. Y., Chen, X., & Zhang, T. (2015). A big data approach for logistics trajectory discovery from RFID-enabled production data. *International Journal of Production Economics*, *165*, 260-272.

[22] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, 98-115.

[23] Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data mining for the internet of things: literature review and challenges. *International Journal of Distributed Sensor Networks*, *2015*, 12.

[24] Nair, C. (2015). Managing big data, from an analog world. *SMT Magazine*, November, 56-62.

[25] Rahim, I. A. (2015). Empowering Urban Innovation Through Convergence of IoT-Big data and Cloud, Forum on Internet of Things: Empowering the New Urban Agenda Geneva, Switzerland, 19 October 2015.

Biography

Dr. Lidong Wang is an Associate Professor at Mississippi Valley State University, USA. He worked at Ohio State University, Mississippi State University, and the University of South Carolina; and conducted projects supported by the Department of Defense, the National Science Foundation, and the National Aeronautics and Space Administration (NASA). His current research interests include Big Data, Internet of Things, cybersecurity, cyber-physical systems, and Industry 4.0, etc. He has published over 70 papers in various journals. He was the President of the Electricity, Electronics & Computer Technology (EECT) Division of the Association of Technology, Management, and Applied Engineering.