IoT-Big Data Software Ecosystems for Smart Cities Sensing: Challenges, Open Issues, and Emerging Solutions

Alexander D. Cartier[1,2], David H. Lee[2], Burak Kantarci[2,3(✉)], and Luca Foschini[4]

[1] Assured Information Security, Inc., Rome, NY 13441, USA
[2] Clarkson University, Potsdam, NY 13676, USA
[3] University of Ottawa, Ottawa, ON K1N 6N5, Canada
burak.kantarci@uOttawa.ca
[4] University of Bologna, 40136 Bologna, Italy

Abstract. The Internet of Things (IoT) architecture primarily consists of massive amounts of heterogeneous objects, equipped with sensing, computing, and communication capabilities to continuously sense the smart cities pulse. The coordinated collection of this data produces relevant scalability and management issues not only in terms of communication but also in storage and computing to process and analyze large amounts of incoming big data streams. In these systems, people also play a pivotal role which includes both social and technical issues, making the design of these solutions a very complex task. This paper overviews the prevalent solutions and architecture design principles in IoT-big data ecosystems for smart cities sensing. Furthermore, we present the needs of IoT-big data software ecosystems by exemplifying existing IoT systems. We also provide useful insights towards future innovation to address open issues and challenges that are identified based on the expected growth of data in the next decade.

Keywords: Big data · Cyber-physical systems · Data analytics Distributed computing · Fog computing · Internet of Things

1 Introduction

Motivated by the ever-increasing amount and value of data gathered and managed by Internet-born companies, such as Google, Facebook, and Amazon, big data software systems, specialized and tailored for the collection, storage, and analysis of those largest data depots, are already a well-established reality. Research efforts in this area not only highlights crucial issues, such as quality of collected data and scalability of the overall sparse and widely distributed big data systems, but also produced significant results, such as in the notable case of NoSQL database solutions (e.g., Cassandra and MongoDB) that, opposed to traditional SQL-based conventional database solutions, support horizontal

scaling by design [1, 2]. Concurrently, advances in communications and device miniaturization enables the Internet of Things (IoT) vision, with smart objects in conjunction with smartphones – that in 2013 outnumbered normal feature phones – acting as sensors deployed over smart cities and equipped with sensing, computing, and communication capabilities.

At the current stage, several research activities in IoT software design focus on overcoming heterogeneity issues in both communication technologies and software Application Programming Interfaces (APIs) for data gathering and management (see also the IoT architectures and applications sidebar). The goal is to collect and analyze incoming big data flows, which are densely available and harvested in urban areas, allowing for a very large-scale fine-grained sensing, by exploiting all personal resources, mobile activities and collaborations. However, the software design principles to develop a new class of IoT-big data solutions defined as software ecosystems are still widely unexplored. For this to become an effective technology, IoT-big data systems for smart cities sensing still have to face several challenges. First, there has been a continuously growing network infrastructure associated with the rise of IoT, generating more data (i.e., high volume) at an exponentially increasing rate (i.e., high velocity). This data rate is overwhelming the conventional means of data management, calling for new methods of collecting, storing, and processing information collected from the Internet of Things. There are several exemplary models and technologies to implement these new concepts in big data management in accordance within the IoT framework. This ecosystem calls for an integrated architecture which combines the functions of Service-Controlled Networking (SCN) through the middleware, orchestrating the service components of the cloud [3]. From the social perspective, smart city sensing, which involves smartphone-enabled sensing, calls for the identification of willing users to participate in sensing campaigns, keeping them involved (e.g., by providing services, entertainment, and rewards), and fostering their participation with active collaboration in data collection. This requires user tasks to operate at specific locations (e.g., taking a picture of a monument, tagging a place, etc.). However, the boundary between social and technical challenges is not clear cut. The technical problem of minimizing the global resource overhead by entrusting a minimal subset of users in a sensing campaign requires analyzing their geo-social profile; to identify and infer which users are most likely to successfully harvest the required data [4].

Focusing on big data, this emerging trend has been clearly recognized by the market. IDC's Digital Universe study in 2012 reports 18% of the United States' digital universe is valuable (158 exabytes in 898 exabytes) when analyzed and tagged properly. In the same report, it is predicted that by 2020, the useful information will grow 17-fold corresponding to 40% of the digital universe forecasted for 2020 (i.e., 2631 exabytes in 6617 exabytes). Another important research discovery suggests that the cloud stored, processed, and transmitted 14% of the digital universe in 2012 whereas cloud-based services will host 37% of the digital universe in 2020. These numbers call for the adoption of big data solutions in IoT where volume and value of the data will be the driving factors in the evolution of

big data software, optimization for real time systems, cybersecurity and forensics for the next decade [5]. These numbers confirm that the challenges with the IoT-big data will keep evolving throughout the next several decades. This article aims to present a collection of methods, and theories, discuss prevalent architecture design principles and identify the required technologies by detecting open issues and challenges, which will provide insights for paving the way towards effective IoT-big data solutions. We provide a survey of remarkable big data solutions for the IoT, then we identify and discuss the open issues and challenges in the IoT-big data ecosystems.

2   IoT Architectures and Applications

The Internet of Things (IoT) pervasively and ubiquitously interconnects billions of devices with sensing, computing and communication capabilities as seen in Fig. 1.

Furthermore, it is crucial to collect, aggregate and correctly represent the data gathered at the sensor network level, where the data will be sent to the next level, namely the middleware. The middleware acts as a layer between the software application layer and hardware layer by parsing the data in order to recognize certain trends or specific patterns to create reusable solutions for frequently encountered problems such as heterogeneity, interoperability, security and dependability. It is worthwhile noting that the majority of middleware solutions currently do not provide the functionality of context-awareness and most solutions are focused on device management. Hence, it will be critical to continue to work on the implementation of context-awareness into the middleware solutions within the IoT [3].

In the IoT-big data ecosystem, the sensing plane consists of the sensors of various types; such as temperature, light, airflow, motion, humidity and several other sensors purposed for various applications including vehicular networks, water quality and e-health monitoring. The role of sensors in this architecture is continuously reporting the sensed data to the data plane through the middleware. The middleware, as mentioned previously is responsible for aggregating data from numerous sensors and presenting it to the data plane for pre-processing and storage. The data plane offers short and long term storage for the aggregated data, and it pre-processes the data for the Cloud platform which provides Data Analytics as a Service [6] where embedded analytics and statistics libraries play a key role [7]. It is worthwhile noting that the data plane can also be implemented within the cloud platform based on the storage-as-a-service concept in cloud systems. The application plane receives software-as-a-service (SaaS) from the cloud platform and interprets the analyzed data in accordance with the desired application e.g., e-health, smart metering, intelligent transportation systems, and so on.

In the corresponding architecture, data is collected via distributed sensors that are uniquely identifiable, localizable and communicable. The collected data goes from the user interaction with the embedded system, up to the local network level and is then stored either on local servers or in the cloud, at which point
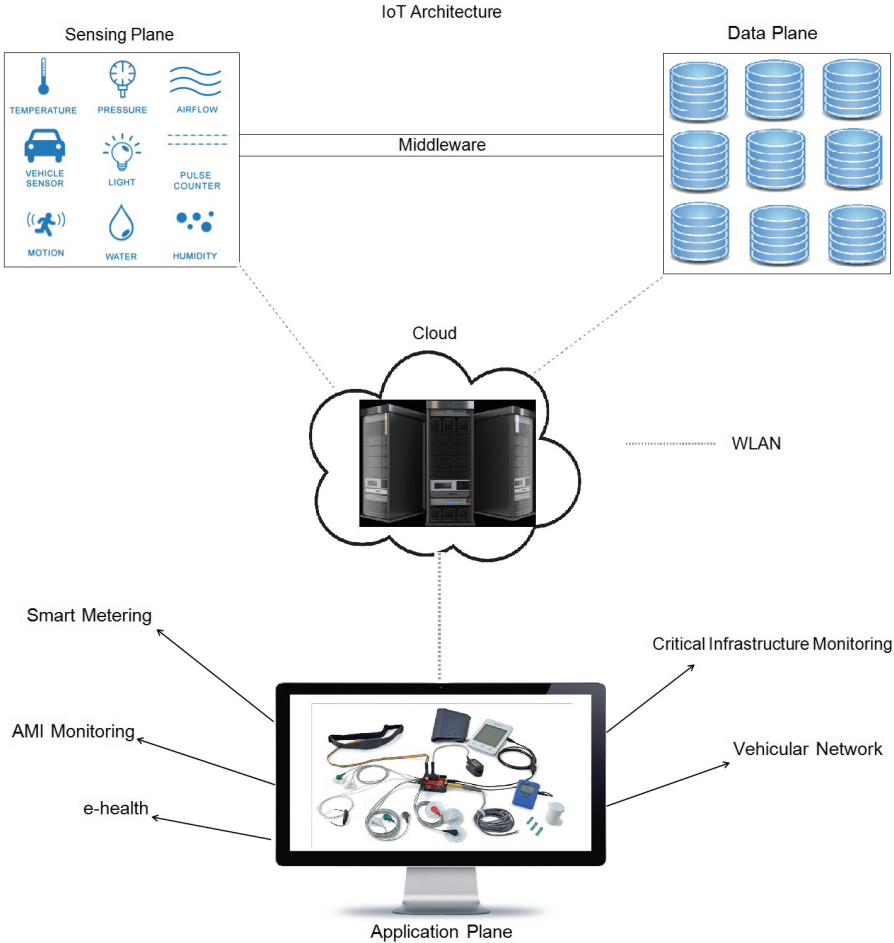
Fig. 1. The IoT-big data ecosystem with three planes that are connected through a cloud platform.

the data is available for a variety of uses. The IoT architecture interconnects sensors, RFID tags, smart phones, and other objects in a scalable manner.

In [8], the requirements of sensing objects driving the integration of cloud computing and IoT are summarized as having huge computing and storage capacity, web-based interfaces for data exchange and integration as well as programming platforms, real-time processing of big data, inter-operability between the sensing objects, cost-efficient, scalable on-demand access to the IT resources, and security and privacy assurance. Therefore, the authors propose deployment, development and management of the IoT applications over the cloud, namely the CloudThings architecture. Recent progress in IoT has not only been made in applications dealing with data analysis, but also in newfound approaches in

structure, storage, and compression. The common goal in the related works is to make the IoT data readily accessible and understandable to the end user.

A motivation of some of the recent progress has been a product of the Smart-Campus project that includes two different scenarios where sensors would be placed to determine the occupation rate of parking lots as well as regulating the temperature through the control of doors leading outside [9]. These services on the network would continuously collect data in real time in order to eventually recognize patterns. The middleware plays a key role in the implementation of this project because of several API's used to send data, set up the configurations for measurement retrieval, and to interact with collected data sets. The responsibility of the middleware is to support the data reception as well as broadcasting the configurations made on the sensors.

3   IoT-Big Data Design Guidelines

Figure 2 illustrates the need for big data management schemes in an IoT-dominated environment. Volume, variety and velocity of the data are driven by the data quality assurance needs, uncertainty in social media accounts, and networked devices being twice as the global population. Consequently, the traditional data warehousing solutions remain with low veracity in the IoT Era.

Starting from big data-related aspects, major trends in the field of big data gathering have an intense focus on the following four areas: velocity, variety, volume and veracity. Velocity denotes a focus on high-speed processing/analysis such as click-streaming and fast database transactions. Variety in the structures of data being collected arises along the lines of Machine-to-Machine (M2M), radio frequency identification (RFID), and different types of sensors. Volume includes currently used services such as social networks, cloud storage, network switches, thermo-metric/atmospheric/motion sensors, and so on. The primary issues that must be considered when focusing on these three subjects are the limitations of nodes' buffer sizes and the maximum acceptable latency in data collection [10]. Finally, veracity is defined as the potential of releasing useful information out of unstructured big data. Indeed, handling of the data through trusted sources improves the veracity of analytics as reported in [7].

The concept of big data has come about with a recent increase in the volume, velocity, variety, and veracity of data collected via various sources but mainly via IoT sensors. A cloud-based eco-system is envisioned to share and trade high-quality data from a vast network of independently managed sensors in real time [1,9]. While there has been considerable research with WSNs, using cloud-based platforms to host sensor networks is one of the biggest challenges yet, and the research regarding this topic has recently started.

This vision introduces a previously unexplored area of research. A few topics which must be addressed in order to find a solution to this challenge are focuses on high-quality data, an efficient collaborative emphasis for sharing/trading data, the need of a markup language that can not only handle the network but can also support data quality and enable domains to access live sensor feeds
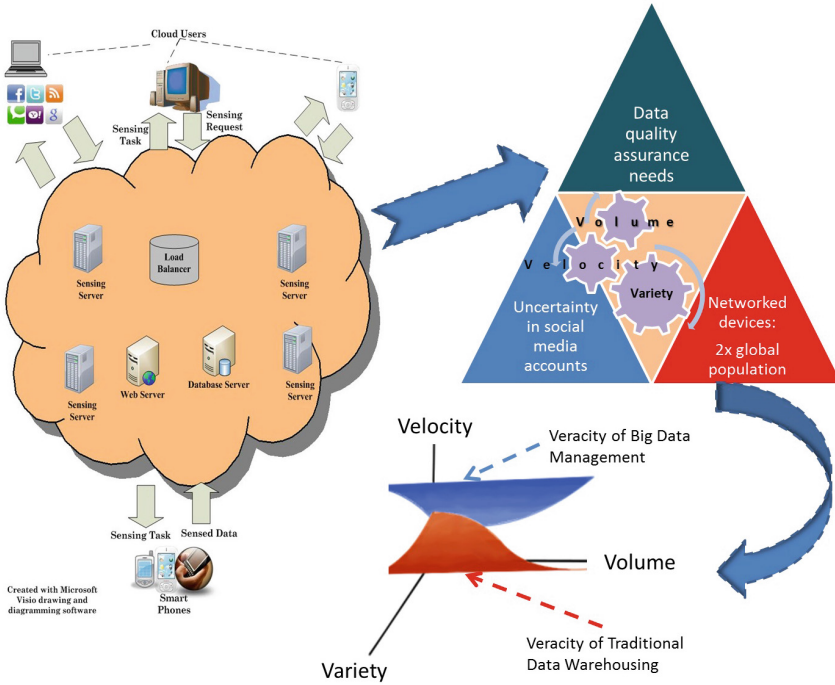
Fig. 2. The rise of the need for big data management in the IoT-dominated environment where majority of the data is collected by connected sensing devices.

as well as historical data [1]. According to this new vision, we propose some main design guidelines and concepts useful to compare existing solutions in IoT-big data literature (see also Table 1).

3.1   State of the Art in Building IoT-Big Data Architectures

Today, IoT-big data systems, such as data collected from the global Flightradar24 flight monitoring system, are handled via software chain architectures. The software chain basically maps processing phases of big data streams to multiple components denoting data generation, intermediate and result stages. As an example, the study in [11] uses the big data stream from a global flight monitoring system and is processed via the Yahoo!S4 framework. The Yahoo!S4 framework is mapped onto five stages, namely the sensor, extractor, parser, formatter and outputter modules.

The sensor module can be implemented as a script which captures unstructured data. The extractor module is responsible for identifying and distinguishing the events within the data streams. On each event, the parser module runs data analytics processes such as filtering, pattern recognition and data mining. The parser module can also be decomposed into multiple layers such as in

the Lambda architecture where Hadoop serves as the batch layer for long term data and Storm serves in the speed layer to manage real time data. Formatter and outputter are responsible for generation of the structured data out of the unstructured data under analysis and maintaining them in a file system of NoSQL database.

In an open IoT system, a similar software chain approach can be adopted. As today's technology is able to enable access to sensor readings through web-based services; the sensor component can obtain the data of the IoT sensors via APIs that enable access to web servers. The Open Geospatial Consortium's (OGC) IoT RESTful API has been built on the OGC Sensor Web Enablement standards in order to interconnect IoT objects, their data and applications over the web via JavaScript Object Notation (JSON) data interchange format. The API can be integrated to the sensor-end of the software chain in order to make various IoT sensors of multiple participants connect to the web servers that are compliant with the OGC standard [12].

3.2 Challenges Experienced in IoT-Big Data Systems

Focusing on IoT-big data systems, we identify some distilled guidelines based on experiences within the ParticipAct sensing project [4], and identify four major categories for design guidelines as support for spatio-temporal queries, minimal overhead on IoT nodes, openness and security, and fast feedback and minimal delay in producing quality-aware sensing data.

Support for Spatio-Temporal Queries. Support for spatio-temporal queries over sensed data is a key factor when considering big data because of all possible sources of where/when data might be pushed from, and it is also important to keep track of sensed data chunks for future uses. First of all, as data is neither temporally nor spatially static in the IoT, storage and scalability of retrieving the data appears as an important issue due to the constant movement of data.

Minimal Overhead on IoT Nodes. By this design guideline, we aim at minimizing energy consumption due to computing and communication at IoT nodes through optimizations of local sensing processes (such as, duty-cycling, employing physical models/verifications, etc.) and, most important, of sensed data transmissions toward the backend (such as, by locally bulking multiple data samples in the same sent packet, coordinating IoT nodes in the same location to avoid useless readings such as in WSN, etc.).

Openness and Security. Sensed data should be stored securely and encrypted to protect it from possible threats. This challenge has been tackled in [13] through the use of a distributed storage system using Shamir's secret sharing as the driving algorithm for both security and storage.

Fast Feedback and Minimal Delay in Producing Quality-Aware Sensing Data. This design guideline derives from the need to associate data with a quality indicator based on the history of data sensed in the past. This requires continuous profiling of sensed data in several different dimensions and grains (such as time, space, weather, season, etc.) by exploiting big data storage to keep all of these profiles ready, thus allowing fast computation for required feedback. Notable efforts within this direction are sensor webs such as IrisNet and SensorWeb. Furthermore, projects like Aurora, Borealis, Cayuga, Stanford Data Stream manager and System have explored many issues associated with stream and event processing comprising the construction of algorithms and techniques for data quality-aware sensor feed discovery service composition [1].

4   Remarkable Big Data Solutions

As big data continues to be researched, there has yet to be a single defining breakthrough when it comes to solutions regarding IoT. This is due to many variables that need to be considered when implementing an idea towards big data in the IoT such as volume, security and storage. While solutions are currently being researched and tested, there have been several instances of progress when dealing with this topic.

4.1   Crowdsensing-Based IoT-Big Data Projects

The features in IoT-big data design guidelines make ParticipAct a complete mobile crowdsourcing platform that encompasses the whole process from data collection, to post-processing, to mining, and is available to the mobile crowdsourcing community as an open-source project [4]. In a related study, after the detailed description of the whole architecture of ParticipAct and its technological stack, some of the use case scenarios are presented. The corresponding scenarios are currently being used to evaluate the potential of mobile crowdsourcing and ParticipAct both qualitatively and quantitatively.

4.2   Smart Environment Projects

The SmartCampus experiment performed on the SophiaTech campus in France [9]. The idea of this project focuses on the final product becoming an open platform for different types of campus members to use the already deployed sensors to build their own services or user defined sensors. Through this project, concepts such as data retrieval and user-defined sensors are implemented into realistic situations where big data and IoT are the focus. Data retrieval is applied in a way where users can pull sensor properties using input filters or just the sensor data itself. User-defined sensors are also introduced in this project as virtual sensors where users can define a specific configuration and store it into database where it can be executed using scripts when its dependencies produce data.

AllJoyn Lambda, a software architecture which integrates the Alljoyn framework into Lambda architecture to enable big data analytics for IoT applications [14]. The proposed architecture adopts the AllJoyn technology that is intended for IoT. However it aims at overcoming the real time processing/storage and management of the data obtained from smart environments by integrating the MongoDB NoSQL database for storage and Apache Storm for real time analysis of the data pushed from the smart objects.

As an emerging smart environment, the IoV concept has evolved from the IoT, and it is presented in [15] where nodes are represented by vehicles and are connected to form a Vehicular Ad Hoc Network (VANET). The biggest challenge in IoV is processing this volume of data and delivering it to its destination, which is done through various relay nodes. This paper analyzes the issue by using Bayesian Coalition Game (BCG) and Learning Automata (LA). The use of BGC trains the LAs to make moves correlating to each node/vehicle performing tasks to make each player in the game safer and more aware. This proposal adopts the Nash Equilibrium concept with respect to the probabilistic belief of players in the coalition game.

4.3   Edge Computing Based Projects

The Available Network Gateways in Edge Location Services (ANGELS) framework appeared as a result of the realization of the complexity of new applications in cyber-physical systems [16]. Services encompassing multiple domains are beginning to come into effect. As astronomical volumes of data collected will begin to require huge computing infrastructures for analysis, ANGELS introduces a framework for fog-computing, which utilizes a key aspect of the IoT field that has been overlooked thus far. The framework focuses on taking into account the ability of resources available prior to the distribution of tasks. Researchers have explored the idea of smart edge devices to perform portions of IoT data analysis where edge devices are low-powered computational nodes such as smart phones and home energy gateways. The proposed architecture consists of servers and commodity computing nodes as well as these smart edge devices as computational resources. This framework of heterogeneous computational nodes includes resources ranging from large server-class systems down to low-powered edge devices forming the basis of the fog computing paradigm. This solution also involves parallel data computation along with capacity based partitioning to accomplish a more streamlined approach to big data management.

4.4   Big Data Stream Analysis Projects for Cyberphysical Systems

The proposal in [17] is tailored for cyber-physical systems, and it presents an online spatiotemporal analysis, which would implement a grid-based single-linkage clustering algorithm over a sliding window. This online time-space efficient method satisfies the velocity demand of big data streams. A large-scale real-world scenario including 300,000 sensors over the course of a year has been established to evaluate the success of the algorithms.

The rising necessity for robust and reliable services is leading to the creation of enormous amount of data, which has the possibility of exceeding the storage capacity of current micro servers. This has led to Big data correlation orchestrator (BigCO) which was implemented in a micro cloud server [18]. In the same study, it is also addressed how multifaceted data could be interrelated and analyzed with 3D modeling. On top of that, a streaming algorithm that extends Ramer-Douglas-Peuker heuristic is presented. This proposed compression algorithm has achieved up to a 99.86% compression of sensor data. With the recognition of consistent growth in the number of wirelessly connected devices, the same study conducts in depth testing dealing with high volumes of data. Their compression method along with the 3D modeling of data assesses the velocity at which they can analyze large quantities of data from a varying pool of sensors. BigCO implementation on a micro cloud server also offers portability to the data collection and analysis mechanism. The overall design of this orchestrator exhibits high veracity throughout the compression, modeling, and the overall BigCO framework.

4.5   Distributed, Secure, Scalable Storage of IoT Data

In [13], a project focusing on secure and scalable IoT storage systems is presented where the security system is derived from Shamir's secret sharing algorithm. However, a major focus is also placed in terms of volume when referring to IoT storage systems. A distributed storage system was designed based on the idea of the algorithm where any sort of incoming data is transformed into scaled shares based on the size of the original files and is inaccessible without the retrieval of all shares. This method considers volume in relation to scalability. This is done through an infrastructure based on a client-peer system where a client takes incoming data; transforms them into scaled shares creating smaller data pieces which can eventually be reassembled to form the original file also taking security into consideration. In terms of performance, this system does not account for the velocity at which the data would be stored and retrieved due to a bottleneck.

4.6   Quality of Data (QoD)-Aware IoT Big Data Projects

In case of continuous retrieval of data through sensor feeds, it is important to focus on the quality of data being pulled down. The project in [1] breaks the idea of the Quality of Data (QoD) down into several aspects, which focuses on availability of sensor feeds, latency, and trustworthiness. These qualities can be used to determine the certain attributes of managing big data in a system where attributes such as trustworthiness of data can be defined as veracity of data or accuracy. Other attributes, such as latency and availability of sensors correlate with the velocity of data being pulled down. The idea behind this project was to create a model for sensor services, where, in order to enable seamless sharing of sensor feeds from various sensors coming from different sources through the cloud. While QoD takes into consideration velocity and veracity of the sensor feeds, they stray from other aspects of big data such as volume and variety.

Though a major focus is placed on the variety of data through the use of heterogeneous sensor feeds, this work does not fully aim to address the variety aspect of big data.

4.7 Spatial Big Data Projects

Due to a gap in the development and applications of integrated information systems for snowmelt flood early warning in water resource management, an integrated system with IoT, geo-informatics (GIS, GPS, etc.), and cloud services have been proposed for the monitoring and simulation of snowmelt flooding [19]. This study resulted in an increase in the effectiveness of decision-making because of the availability of data and the integrated system to analyze all of the data in an efficient enough manner to make a difference when it comes to split-second decisions. This proposal goes off a popular practice in the field, which utilizes environmental tracking and analysis. The architecture of this system collects a wide variety of data from an assortment of information acquisition facilities. The collected data calls for a storage facility that can handle large volumes of information which this system architecture also takes into account. Then the proposals computing and analysis facility as well as the network and software used in this system accounts for high velocity collection and analysis of the data without a loss in veracity through the whole process.

Projects specifically addressing spatial big data exhibits new challenges where performance on large amounts of measurements is associated to specific locations and the instance of time when they were conducted. In addition, the paradigm of wireless networks data analytics varies from the classical data-mining paradigm as it poses different challenges in scalability and computational time. While relational data requires a linear time scale for computation for classification and prediction, spatial data requires a cubic time scale which causes problems for scalability both in terms of volume, calculation period and velocity. Although it remains as a distinct category from relational big data, the research in [20] has shown that existing parallel processing and computational framework algorithms are powerful tools for implementing spatial processing frameworks but the proper architecture for these tools are still being researched.

Dealing with spatial big data is considered one of the key challenges for the development of future wireless networking applications in terms of big data. Some underlying issues with this topic are that it currently requires a high level of specialized knowledge in order to design and implement systems for processing spatial big data. In order for this technology to grow and expand, it requires a wider use of context and development of systems such that non-experts in the field are able to build various applications using this technology. Therefore, a more common language for reasoning and computational inference solutions is necessary for development of these systems.

Table 1. Summary and comparison of the surveyed solutions

Application	Volume	Velocity	Variety	Veracity	Spatio-temp. queries	Min. overhead	Openness-security	Quality-aware sensing
Participact [4]	✓	✓	✓	✓	✓	✓	✓	✓
Smart campus [9]	✓	✓	✓	×	✓	×	×	×
Alljoyn-Lambda [14]	✓	✓	✓	×	✓	×	×	✓
IoV [15]	✓	×	×	✓	×	✓	×	✓
ANGELS [16]	✓	×	✓	×	×	✓	×	✓
CPS [17]	✓	✓	×	×	✓	✓	×	×
BigCo [18]	✓	✓	✓	✓	×	✓	×	×
Dist. storage [13]	✓	×	×	×	×	×	✓	×
Quality of data [1]	×	✓	×	✓	×	×	✓	✓
Flood warning [19]	✓	✓	✓	✓	✓	✓	×	×
Spatial data [20]	✓	✓	×	×	✓	×	×	✓

5   Summary

This paper has introduced the challenges in the IoT-big data ecosystem, and overviewed recent applications as summarized in Table 1. A majority of these applications focuses on two or three of the four Vs typical of big data systems, yet solutions addressing all dimensions are emergent. Veracity is the most neglected dimension in related work; hence trustworthiness assessment modules are emergent in software architectures that are proposed for IoT data management. Focusing on specific IoT-big data challenges, availability of tools and libraries for embedded data analytics are critical for use in the development of middleware solutions for the IoT-big data ecosystem.

All these challenges, pointed out by the related work, will impact the handling of the IoT data which is expected to contribute to the majority of the data accumulation in the near future. NoSQL-based solutions are feasible to overcome the storage challenges of big volumes of data while realtime analytics solutions such as Storm or Spark suit well with the IoT stream data of high velocity.

6   Open Issues and Challenges

As the data will be scaled out to higher volumes, varieties and velocities with the wide adoption of connected devices, IoT and big data will be two inseparable phenomena of the future. Major innovation should be on the analytics software architecture that can handle analytics on long term and real time data. Despite the availability of architectures, like Lambda, that address this issue, optimization software is required for real time systems. OGC's standardization efforts for accessing IoT data is invaluable as heterogeneity of big data pushed by the IoT objects can be handled. More importantly, we expect that even by 2020 less than half of the digital universe will consist of useful data because the majority of the data will still be unstructured and untagged. Therefore, collection of the data, proper tagging and structuring to improve the value of the IoT

data is critical. Furthermore, how to secure the data collected from IoT devices and ensure privacy of personalized devices is a key problem. In fact, cloud-based storage, processing and transmission of the data will be reaching 40% of the digital universe, introducing security as a service in cloud analytics as an emergent issue. Finally, IoT sensors will push IoT data continuously, and typically in raw form to be processed so to produce value. Hence, development of scalable and analytics-backed visualization mechanisms for long term data is also important to prevent data overloads for the IoT-big data systems.

References

1. Ramaswamy, L., Lawson, V., Gogineni, S.V.: Towards a quality-centric big data architecture for federated sensor services. In: IEEE International Congress on Big Data, pp. 86–93 (2013)
2. Gorton, I., Klein, J.: Distribution, data, deployment: software architecture convergence in big data systems. IEEE Softw. **32**(3), 78–85 (2015)
3. Sowe, S., Dong, M., Kimata, T., Zettsu, K.: Managing heterogeneous sensor data on a big data platform: IoT services for data-intensive science. In: IEEE 38th International Computers, Software and Applications Conference Workshops, pp. 295–300 (2014)
4. Cardone, G., Corradi, A., Foschini, L., Ianniello, R.: ParticipAct: a large-scale crowdsensing platform. IEEE Trans. Emerg. Top. Comput. **4**(1), 21–32 (2015)
5. Gantz, J., Reinsel, D: IDC: the digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east (2012). https://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf
6. Talia, D.: Clouds for scalable big data analytics. IEEE Comput. **46**(5), 98–101 (2013)
7. Louridas, P., Ebert, C.: Embedded analytics and statistics for big data. IEEE Softw. **30**(6), 33–39 (2013)
8. Zhou, J., Leppanen, T, Harjula, E., Ylianttila, M., Ojala, T., Yu C., Jin, H., Yang, L.T.: CloudThings: a common architecture for integrating the Internet of Things with cloud computing. In: IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 651–657 (2013)
9. Cecchinel, C., Jimenez, M., Mosser, S., Riveill, M.: An architecture to support the collection of big data in the Internet of Things. In: IEEE World Congress on Services (SERVICES), pp. 442–449 (2014)
10. Takashi, D., Nishiyama, H., Kato, N., Miura, R.: Toward energy efficient big data gathering in densely distributed sensor networks. IEEE Trans. Emerg. Top. Comput. **2**(3), 388–397 (2014)
11. Xhafa, F., Naranjo, V., Caballe, S., Barolli, L.: A software chain approach to big data stream processing and analytics. In: Ninth International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS), pp. 179–186 (2015)
12. The OGC (Open Geospatial Consortium) SensorThings API Standard. https://ogc-iot.github.io/ogc-iot-api/index.html
13. Jiang, H., Chen, S., Li, K.-C., Jeong, Y.-S.: A secure and scalable storage system for aggregate data in IoT. Future Gener. Comput. Syst. **49**, 133–141 (2015)

14. Villari, M., Celesti, A., Fazio, M., Puliafito, A.: Alljoyn Lambda: an architecture for the management of smart environments in IoT. In: International Conference on Smart Computing Workshops, pp. 9–14 (2014)
15. Kumar, N., Misra, S., Rodrigues, J., Obaidat, M.: Coalition games for spatio-temporal big data in internet of vehicles environment: a comparative analysis. IEEE IoT J. **2**(4), 310–320 (2015)
16. Mukherjee, A., Paul, H.S., Dey, S., Banerjee, A.: Angels for distributed analytics in IoT. In: IEEE World Forum on Internet of Things (WF-IoT), pp. 565–570 (2014)
17. Fu, Z., Almgren, M., Landsiedel, O., Papatriantafilou, M.: Online temporal-spatial analysis for detection of critical events in cyber-physical systems. In: IEEE International Conference on Big Data, pp. 129–134 (2014)
18. Mozumdar, M., Shahbazian, A., Ton, N-Q.: A big data correlation orchestrator for Internet of Things. In: IEEE World Forum on Internet of Things (WF-IoT), pp. 304–308 (2014)
19. Fang, S., Xu, L., Zhu, Y., Liu, Y., Liu, Z., Pei, H., Yan, J., Zhang, H.: An integrated information system for snowmelt flood early-warning based on Internet of Things. Inf. Syst. Front. **17**(2), 321–335 (2015)
20. Jardak, C., Mahonen, P., Riihijarvi, J.: Spatial big data and wireless networks: experiences, applications, and research challenges. IEEE Netw. **28**(4), 26–31 (2014)