4th International Conference on
Combinatorics, Cryptography, Computer Science and Computing
November 20-21, 2019 - Iran University of Science & Technology

# Data Analysis, Analytics in Internet of Things and BigData

**Leila Damghani [1], Hamidreza Damghani [2], Heliasadat Hosseinian [3], Hossein Mohammadnezhad Shourkaei [4]**

[1] Faculty of Computer Engineering- Artificial intelligence, Kharazmi University, Tehran, Iran.

[2] Faculty of Mechanics, Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

[3] Faculty of Mechanics, Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

[4] Faculty of Mechanics, Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

**ABSTRACT**

The Internet-of-Things (IoT) is gradually being established as the new computing paradigm, which is bound to change the ways of our everyday working and living. IoT emphasizes the interconnection of virtually all types of physical objects (e.g., cell phones, wearables, smart meters, sensors, coffee machines and more) towards enabling them to exchange data and services among themselves, while also interacting with humans as well. Few years following the introduction of the IoT concept, significant hype was generated as a result of the proliferating number of IoT-enabled devices, which (according to many projections) are expected to amount to several billion in the next years. During recent years, this hype has been turning to reality, as a wave of IoT applications with significant social and economic has been emerging. Data analytics is the process of deriving knowledge from data, generating value like actionable insights from them. This article reviews work in the IoT and big data analytics from the perspective of their utility in creating efficient, effective and innovative applications and services for a wide spectrum of domains. We review the broad vision for the IoT as it is shaped in various communities, examine the application of data analytics across IoT domains, provide a categorization of analytic approaches and propose a layered taxonomy from IoT data to analytics. IoT data analysis is an integral element of any non-trivial IoT system. Nevertheless, IoT analytics are still in their infancy, as IoT data still remain largely unexploited.

**KEYWORDS:** Internet of Things; Analytics; BigData; Data mining; Database management.

# 1 INTRODUCTION

The internet-of-things (IoT) paradigm represents one of the next evolutionary steps in Internet-based computing, which is already having a positive impact in a large number of application domains including smart cities, sustainable living, healthcare, manufacturing and more. IoT analytics refers to the analysis of data from multiple IoT data sources, including sensors, actuators, smart devices, and other internet connected objects. The collection and analysis of data streams from IoT sources is nowadays considered a key element of the IoT's disruptive power, as well as a prerequisite to realizing IoT's hyped market potential. Indeed, according to a recent report by McKinsey [1], less than 1% of IoT data is currently used, which is a serious setback to maximizing IoT's business value. For example, most IoT analytics applications are nowadays used for anomaly detection and control rather than for optimization and prediction, which are the applications that will provide the greatest business value in the coming years.

## *1-1 IoT Data and BigData*

The rise of future internet technologies, including cloud computing and BigData analytics, enables the wider deployment and use of sophisticated IoT analytics applications, beyond simple sensor processing applications. It is, therefore, no accident that IoT technologies are converging with cloud computing and BigData analytics technologies towards creating and deploying advanced applications that process IoT streams [2]. The integration of IoT data streams within cloud computing infrastructures enables IoT analytics applications to benefit from the capacity, performance, and scalability of cloud computing infrastructures. In several cases, IoT analytics applications are also integrated with edge computing infrastructures, which decentralize processing of IoT data streams at the very edge of the network, while transferring only selected IoT data from the edge devices to the cloud. Therefore, it is very common to deploy IoT analytics applications within edge and/or cloud computing infrastructures [3]. IoT data are essentially BigData since they feature several of the Vs of BigData, including (Figure 1):

- ***Volume:*** IoT data sources (such as sensors) produce in most cases very large volumes of data, which typically exceed the storage and processing capabilities of conventional database systems.
- ***Velocity:*** IoT data streams have commonly very high ingestion rates, as they are produced continually, in very high frequencies and in several times in very short timescales.
- ***Variety:*** Due to the large diversity of IoT devices, IoT data sources can be very heterogeneous both in terms of semantics and data formats.
- ***Veracity:*** IoT data are a classical example of noise data, which are characterized by uncertainty [4].

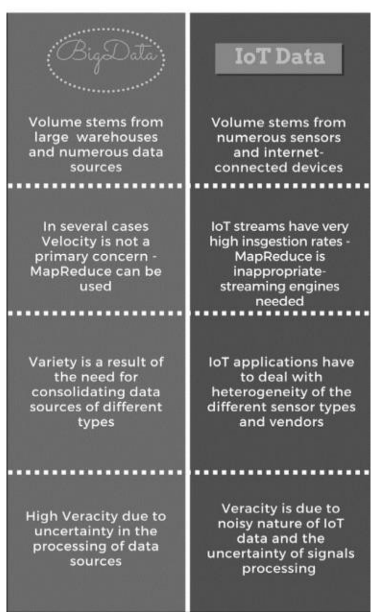

Figure.1. BigData Vs IoT (Big) Data.

The IoT is a complex network of things and people that are seamlessly connected through the Internet. Literally [5], anything that can be connected will be connected and communicating using wireless sensors and RFID tags [6]. And once connected, these connected "things" can send data and interact with other things and people–all in real time. Sensors alone can generate massive volumes of data–unstructured data that can be classified, organized, analyzed and harnessed as part of a super-fast feedback loop to support accurate, automated decisions and actions. This feedback loop also allows you to see, monitor and control things remotely through the Internet, as well as continuously improve operations. It also ensures you're no longer in the dark about what's really going. With real-time alerts and insights, you can take intelligent, instant action to address issues and achieve goals [7].

Data-driven companies are already using machine-generated data from the IoT to enhance customer service, generate more revenue from new products and services, optimize operations, and feed more data into existing analytics efforts. They are also using it to:

• Move from selling products to selling end-to-end services
• Build new and innovative products
• Reduce system downtime and identify and resolve network bottlenecks
• Improve customer experience
• Increase the productivity of existing operations and infrastructure
• Make smarter decisions regarding future infrastructure investments
• Predict and improve mean-time-to-failure for machinery and other capital-intensive assets

### *1-2 Data analyzing*

Collecting, preparing and analyzing all of this streaming, fragmented data is no small task. The data volumes can double every few months, and the data itself is complex – often in hundreds of different semi-structured and unstructured formats. And most importantly, to gain the big insights from streaming sensor data, you need to be able to manage and analyze all of your structured and unstructured data together, all at once. If you can't, the potential insights you can generate are significantly limited [1, 8]. This means moving beyond the limitations of traditional enterprise data warehouses (EDWs) and business intelligence (BI) software. EDWs can't handle unstructured data, so IT has to try to force structure upon unstructured data before business users can analyze it. The problem is, this takes too much time – and any attempt to structure unstructured data in tables limits its potential value as a source of insight.

This is where big data analytics comes into play. It allows you to:

• Combine, integrate, and analyze all of your data at once – structured, semi-structured, and unstructured–regardless of source, type, size, or format

• Quickly and affordably scale to huge volumes of data and analyze them for insights.

## 2 CHALLENGES OF IOT ANALYTICS APPLICATIONS

***2-1 The heterogeneity of IoT data streams****:*IoT data streams tend to be multi-modal and heterogeneous in terms of their formats, semantics, and velocities. Hence, IoT analytics applications expose typically variety and veracity. BigData technologies provide the means for dealing with this heterogeneity in the scope of operationalized applications. However, accessing IoT data sources (including sensors and other types of internet-connected devices) requires drivers and connectors [4, 9], beyond what is typically deployed in transactional BigData applications (e.g., database drivers). Furthermore, dealing with semantic interoperability of diverse data streams requires techniques beyond the (syntactic) homogenization of data formats.

***2-2 The varying data quality****:* Several IoT streams are noisy and incomplete, which creates uncertainty in the scope of IoT analytics applications. Statistical and probabilistic approaches must be therefore employed in order to take into account the noisy nature of IoT data streams, especially in cases where they stem from unreliable sensors [3, 10, 11]. Also, different IoT data streams can be typically associated with different reliability, which should be considered in the scope of their integration in IoT analytics applications [2, 12].

***2-3 The real-time nature of IoT datasets**:* IoT streams feature high velocities and for several application must be processed nearly in real-time. Hence, IoT analytics can greatly benefit from data streaming platforms, which are part of the BigData ecosystem. IoT devices (e.g., sensors) provide typically high-velocity data, which however can be in several cases controlled by focusing only on changes in data patterns and reports, rather than dealing with all the observations that stem from a given sensor.

***2-4 The time and location dependencies of IoT streams**:* IoT data come with temporal and spatial information, which is directly associated with their business value in a given application context. Hence, IoT analytics applications must in several cases process data in a timely fashion and from proper locations. Cloud computing techniques (including edge computing architectures) can greatly facilitate timely processing of information from given locations in the scope of large scale deployments. Note also that the spatial and temporal dimensions of IoT data can serve as a basis for dynamically selecting and filtering streams towards analytics applications for certain timelines and locations.

***2-5 Privacy and security sensitivity**:* IoT data are typically associated with stringent security requirements and privacy sensitivities, especially in the case of IoT applications that involve the collection and processing of personal data. Hence, IoT analytics need to be supported by privacy preservation techniques, such as the anonymization of personal data, as well as techniques for encrypted and secure data storage.

***2-6 Data bias**:* As in the majority of data mining problems, IoT datasets can lead to biased processing and hence a thorough understanding and scrutiny of both training and test datasets are required prior to their operationalized deployment. To this end, classical data mining techniques can also be applied in the IoT case. Note that the specification and deployment of IoT analytics systems entail techniques similar to those deployed in classical data mining problems, including the understanding of the data, the preparation of the data, the testing of data mining techniques and ultimately the development and deployment of a system that yields the desired performance and efficiency [13].

## 3 IOT ANALYTICS LIFECYCLE AND TECHNIQUES

The IoT analytics lifecycle comprises the phases of data collection, analysis, and reuse. In particular:

• **Phase 1 – IoT Data Collection**: As part of this phase IoT data are collected and enriched with the proper contextual metadata, such as location information and timestamps. Moreover, the data are validated in terms of their format and source of origin. Also, they are validated in terms of their integrity, accuracy, and consistency. Hence, this phase addresses several IoT analytics challenges, such as the need to ensure consistency and quality. Note that IoT data collection presents several peculiarities, when compared to traditional data consolidation of distributed data sources, such as the need to deal with heterogeneous IoT streams.

• **Phase 2 – IoT Data Analysis**: This phase deals with the structuring, storage and ultimate analysis of IoT data streams. The latter analysis involves the employment of data mining and machine learning techniques such as classification, clustering and rules mining. These techniques are typically used to transform IoT data to actionable knowledge.

• **Phase 3– IoT Data Deployment, Operationalization, and Reuse**:

As part of this phase, the IoT analytics techniques identified in the previous steps are actually deployed, thus becoming operational. This phase ensures also the visualization of the IoT data/knowledge according to the needs of the application. Moreover, it enables the reuse of IoT knowledge and datasets across different applications [3]. The tasks outlined in the above-listed phases are supported by a range of data management and analysis disciplines, including:

• **IoT middleware and interoperability technologies**, which provide the means for collecting, structuring and unifying IoT data streams, thus addressing the variety and veracity challenges of IoT data.

• **Statistics**, which provide the theory for testing hypotheses about various insights stemming from IoT data.

• **Machine learning**, which enables the implementation of learning agents based on IoT data mining. Machine learning includes several heuristic techniques. The practical cases studies in the second part of the book make use of various machine learning schemes.

• **Data mining and Knowledge Discovery**, which combines theory and heuristics towards extracting knowledge. To this end, data cleaning, learning, and visualization might be also employed.

• **Database management systems**, including Relational Database Management Systems (RDMS), NoSQL databases, BigData databases (such as the HDFS (Hadoop Distributed File System), which provide the means for data persistence and management. Most of the practical examples and case studies presented in the book make use of some sort of database management systems in order to persist and manage the data [8].

• **Data streams management systems**, which handle transient streams, including continuous queries, while being able to handle data with very high ingestion rates, including streams featuring unpredictable arrival times and characteristics. IoT streaming systems are also supported by scalable, distributed data management systems [4].

## 4 CONCULATION

In this paper has defined the scope of IoT analytics and presented related technologies. It has also outlined the close affiliation of IoT analytics with cloud computing and BigData techniques. Furthermore, it has presented the main challenges of IoT analytics applications, which stem primarily from the unique characteristics and nature of IoT data. The rest of the book is destined to present technology solutions to these challenges, along with practical applications and case studies, which make use of such solutions. The presented solutions build in several cases over state-of-the-art IoT, cloud computing and BigData solutions, given that the integration of these technologies tends to become a norm for the variety of IoT analytics applications.

## REFERENCES

[1] Manyika, J., Chui, M., Bisson, P., Woetzel, J., Dobbs, R. Bughin, J., Aharon, D. Unlocking the Potential of the Internet of Things. McKinsey Global Institute, June (2015).

[2] J. Soldatos, et al. “IoT analytics: Collect, Process, Analyze, and Present Massive Amounts of Operational data – Research and Innovation Challenges” Chapter 7 in Book, “Building the Hyperconnected Society –IoT Research and Innovation Value Chains, Ecosystems and Markets”, IERC Cluster Book 2015, River Publishers.

[3] H. Hosseinian and H. Damghani, “Smart home energy management, using IoT system,” in *5th Conference on Knowledge Based Engineering and Innovation (KBEI), IEEE*, 2019.

[4] H. Damghani, H. Hosseinian, and L. Damghani, “Overview of security aspects for LTE and LTE-A networks,” in *14th Media Technology Conference,* Islamic Republic of Iran Broadcasting, Tehran, Iran, December 2017.

[5] H. Damghani, “RFID Technology: Benefits and Applications,” *Monthly Journal of Transportation and Development*, vol. 36, pp. 70–73, August 2010.
[6] H. Damghani, H. Hosseinian, and L. Damghani, “Investigating attacks to improve security and privacy in RFID systems using the security bit method,” in *5th Conference on Knowledge Based Engineering and Innovation (KBEI), IEEE,* 2019.
[7] H. Hosseinian, H. Damghani, L. Damghani, G. Nezam and H. Hosseinian, “Home appliances energy management based on the IoT system,” *The International Journal of Nonlinear Analysis and Applications,* 2019.
[8] H. Hosseinian, H. Damghani, L. Damghani, E. Kouhi and M. Kouhi, “Asia’s Cities: The way to Going Smart,” *The International Journal of Nonlinear Analysis and Applications,* 2019.
[9] H. Damghani and L. Damghani, “Security improvement of Common Scrambling Algorithm (CSA) using the encryption extension technique,” in *16th Iran Media Technology Exhibition & Conference (IMTEC2019),* IEEE, November 2019.
[10] H. Damghani, H. Hosseinian, and L. Damghani, “Bitcoin through Amateur Radio,” in *4th Conference on Technology In Electrical and Computer Engineering (ETECH2019),* 2019.
[11] H. Damghani, H. Hosseinian, and L. Damghani, “Cryptography review in IoT,” in *4th Conference on Technology In Electrical and Computer Engineering (ETECH2019),* 2019.
[12] Tadapaneni, N. R. (2019). Role of Fog Computing in the Internet of Things. International Journal of Scientific Research and Engineering Trends.
[13] H. Damghani, H. Hosseinian, L. Damghani and F. Faghihi, “Standardization for the Internet of Things,” in *4th International Conference on Combinatorics, Cryptography, Computer Science and Computation*, November 2019.
[14] H. Damghani, L. Damghani, H. Hosseinian and R. Sharifi, “Classification of Attacks on IoT,” in *4th International Conference on Combinatorics, Cryptography, Computer Science and Computation,* November 2019.
[15] Prasad, K., & Ramchander, K. M. Layer Architecture Of Privacy Preserving Cloud Storage Using Fog Computing.
[16] Ambrose, T. V., Chidhambaran, R., & Mohan, M. C. Technological changes: Recent trends and challenges in banking sector.