ECONTECHMOD. AN INTERNATIONAL QUARTERLY JOURNAL – 2018, Vol. 7, No. 3, 7–12

Fog computing and Big data in projects of class smart city

O. Duda[1], N. Kunanets[2], O. Matsiuk[1], V. Pasichnyk[2], A. Rzheuskyi[2]

[1]Ternopil Ivan Puluj National Technical University;
e-mail: oleksij.duda@gmail.com; oleksandr.matsiuk@gmail.com
[2]Lviv Polytechnic National University;
e-mail: vpasichnyk@gmail.com, nek.lviv@gmail.com, nek.lviv@gmail.com

Received July 06.2018: accepted October 20.2018

Abstract. The possibilities of using the information technology of foggy computing, which implements the processes of selecting primary messages from sensory nodes and their processing, and further transferring the results of primary computing to server environments based on information technology platforms of cloud type are analyzed in the article. It is noted that both of the approaches – mist and cloud computing – can be effectively used in a wide range of applications, in particular those that are used in information and technology complexes of "smart cities". The system of parameters which distinguishes a separate class of information technologies called Big data is analyzed in the paper. The analysis made it possible to fix 10 basic parameters of the so-called 10 v, with the help of which a separate class of information technologies is allocated. Big data technologies, along with the technologies of foggy and cloud computing are elements of a powerful information technology platform that allows us to solve a wide range of problems for smart cities. The authors illustrated the system-technological connections of these classes of information technologies and analyzed the possibilities of their use in the context of implementation of the information technology project "Ternopil Smart city".

Key words: cloud technology, foggy computing, Big data, information technology platform, smart city.

INTRODUCTION

The current pace of technological progress and globalization, due in particular to the total introduction of information and communication technologies such as social networks, e-Commerce, Internet devices (IoT), data sensor, genomics and cloud computing, generate an exponential growth in the volume of data sets, which significantly exceeds the technological capabilities of their processing. IDC [1] predicts that by 2020, the amount of information collections presented in digital formats will reach 40 zebabytes (ZB) which is 40 trillion gigabytes of data, while the number of transactions in B2C and B2B class technologies will reach 450 billion a day, which in its turn leads to the formation of a new class of tasks aimed at processing a large number of data called "Big data", generates the need to increase computing power and volumes of repositories, development of means for their analytical processing.

STATUS OF THE PROBLEM STUDY

The analysis of the use of "foggy" technologies in various subject areas was conducted by many authors [2, 3, 4]. The work [5, 6] describes the characteristics of "Big data", in the interpretation of the specialists of Gartner [7], who entered the professional circulation as the concept of "3V": volume, variety and velocity.

The international data corporation (IDC) defines "Big data" information technology as "a new generation of technologies and architectures designed to efficiently process a large variety of data, allowing for high-speed data collection, identification and analysis" [8] and notes that "Big data" is not fully characterized by the concept of "3V", which requires the expansion of the list of features to "4V" by adding as characteristic of "Big data" the property of which is value.

Since, using the concept of "Big data" it should take into account the investment costs for its implementation, together with a decrease in operating costs and improving the quality of service.

The initial quality and cost of data will be low, but the results of analytical processing of data significantly increase their value. The data scientists should assess the benefits and costs of collecting or creating applications based on "Big data", choosing high-quality data sources, and implementing analytical methods capable of increasing the added value of information tuples of data and knowledge.

At the same time, there was no research on the possibility of using these technologies for building information technology platforms of smart city that would allow analytical processing of Big data accumulated on the basis of "fog" technologies.

FOG COMPUTING AS AN EFFECTIVE MEAN OF IMPLEMENTING IOT TECHNOLOGY IN "SMART CITIES"

The concept of "fog computing" allows us to move computing processes from cloud environments to network resources [2], which are as close as possible to data sources located at the lower (sensory) level. "Fog computing" is essentially a virtualized information and technology platform that allows to provide services for computing, data storage and networking between basic IoT devices and cloud platforms [3]. "Fog computing" is an architecture that is designed to provide information exchange and network management processes [4], in particular, "fog" nodes receive source data from the sensory level, perform their preliminary processing and send them via Internet to the cloud information technology platform of "smart city". "Fog" nodes, depending on the context, can act both as clients and as servers. In terms of sensor devices, "fog" nodes are application servers that receive data and are considered as access points because they create private networks for communication between sensory and "fog" levels. On the other hand, the "fog" nodes are clients of cloud services and applications, which, are connected over the Internet, for the purpose of sending data from the sensory level.

Fig. 1 shows the general architecture of "fog computing", in which the presentation of various "fog" services is provided by the use of resources of IoT devices and primary networks, including wireless sensor networks (WSNs), virtual sensor networks (VSN)), vehicular ad hoc network (VANETs) and personal area networks (PANs) [9].

Fig. 2 shows the main levels of the "fog computing" architecture [10]. The physical level and virtualization level is the lowest level to which devices ("things") that generate data and are able to connect to a local network or the Internet. It contains nodes, devices, physical and virtual sensors, physical and virtual sensor networks, data communication, etc. This kind of nodes are managed according to the requirements of services and characteristics of one or another node. At the monitoring level, the functioning of nodes and networks is controlled in accordance with the tasks and available computing power. At this level, the order and time of the tasks are determined and the power consumption of network nodes is controlled. The level of previous processing is responsible for data management. At this level, a preliminary analysis of data is performed on the basis of which the excess values are cropped and filtered.

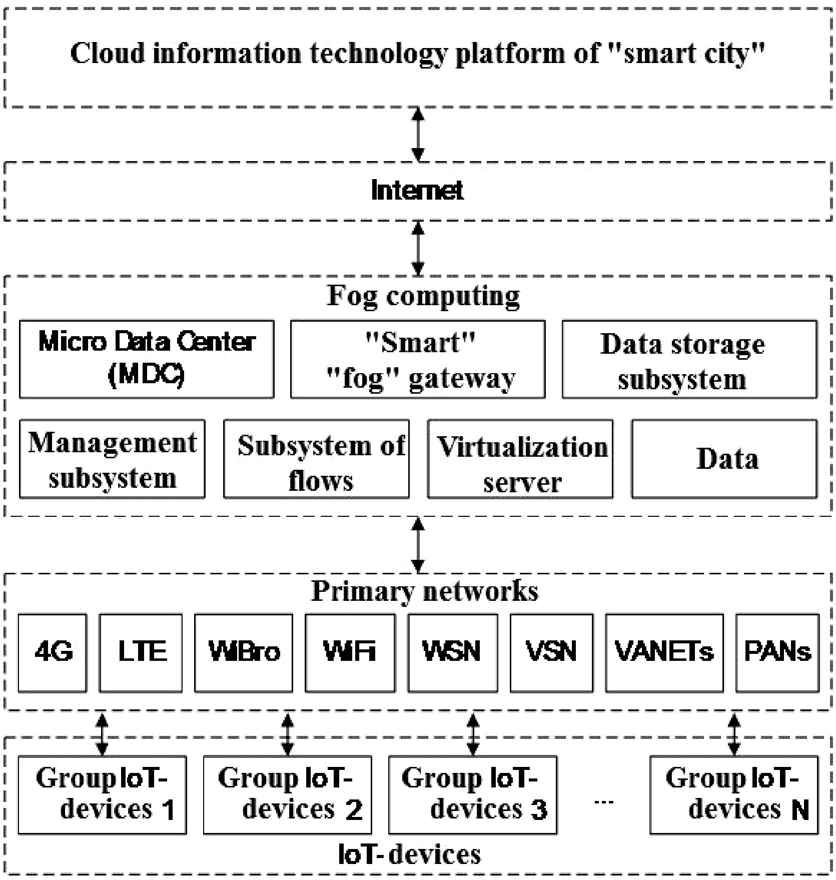

Fig. 1. Generalized architecture of "fog computing"

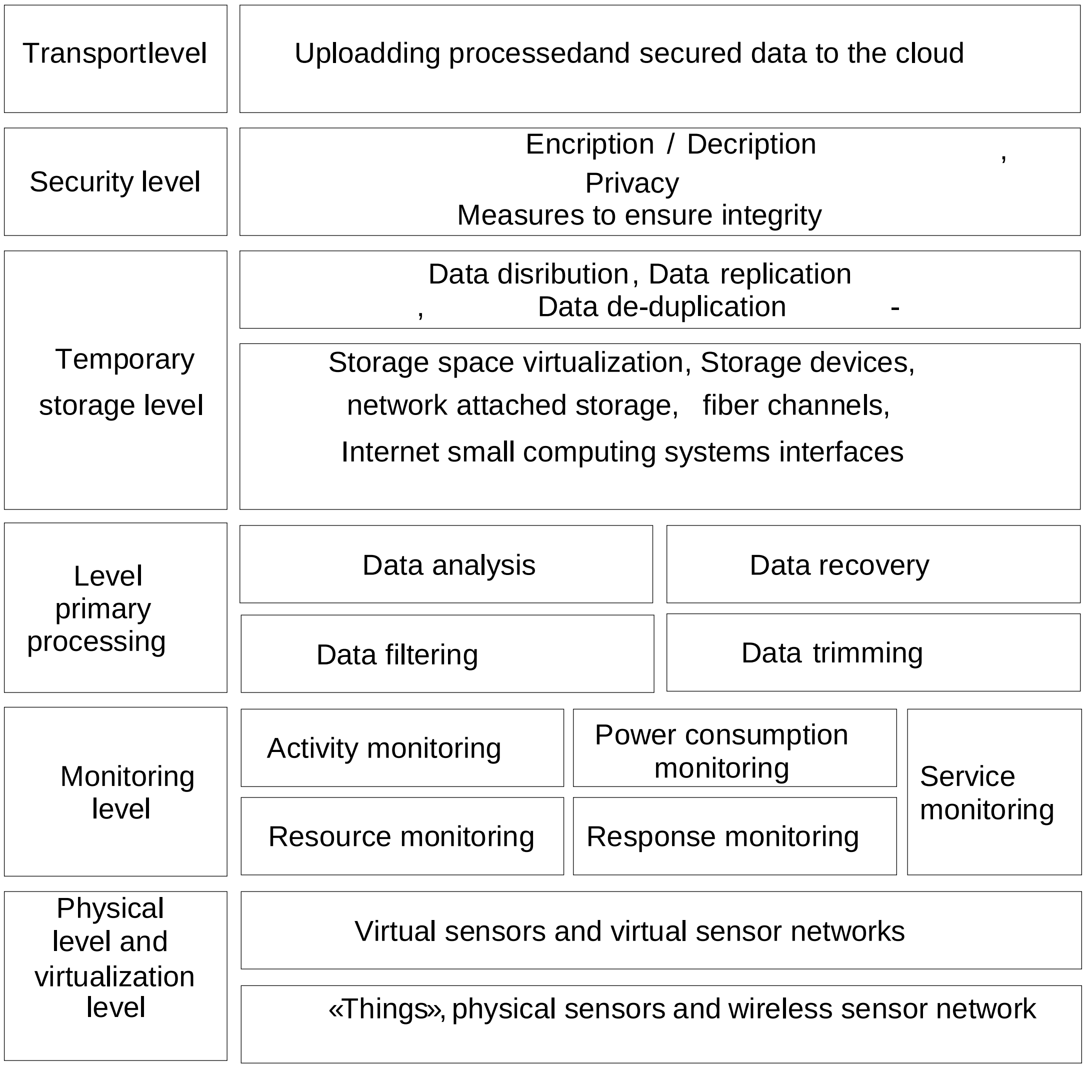

Fig. 2. The main levels of “fog computing” architecture [11]

The level of temporary storage is responsible for storing “fog” data that are at the specified level in the status of temporary. For long-term storage of data sets, the cloud-based information and technology platform of the “smart city”, which has much more available resources is used. The implementation of daily processes of the “smart city” by means of attracting the “smartness” services and resources of the relevant services stipulates the formation of powerful information tuples containing extensive personal data of the residents and guests of the city and that they must adhere to the relevant rules of safety and confidentiality. The level of protection of the “fog” information technology platform ensures the implementation of appropriate functions before sending data through public or vulnerable channels. To transfer data to the cloud information technology platform of the “smart city”, a transport layer is used that reduces the load on the basic network and allows the equipment to provide the enhanced services more quickly. Since messages among “fog” nodes and the “cloud” information technology platform of the “smart city” are transmitted over the Internet, it is important that “fog” nodes use protocols with implemented authorization functions and data integrity. In this case, it is advisable to conduct the certification of each “fog” node by means located in “cloud” environment. Usually, “fog” resources are placed among IoT nodes and “cloud “information technology platform of the “smart city” [12], allowing the creation of rationally organized, more specialized and context-sensitive services [13] which, thanks to the proximity to the basic sensors, provide a reduction in delay time and high-quality streaming of mobile network nodes [14], including moving vehicles and road infrastructure (for example, mobile sensors, roads, railways and vehicles) [15], that in turn provides better integration of network and “cloud” resources. “Fog” computing make it possible to facilitate the implementation of the processes of coordination of different types of urban data generated by heterogeneous sensors integrated into the “smart city” environment by accelerating the recoding and reformatting of data sets at the local level. Therefore, Big data is accumulated in the “cloud”.

BIG DATA: CONCEPTS, TERMS, AND PARAMETERIZATION

"Big data" is a term used to indicate the exponential increase in the amount of data that is difficult or impossible to store, process and analyze using the tools inherent in the technology of traditional databases. " Big data" is characterized by weak structuring and requires the use of significant computing power for the implementation of identifying useful information sets and knowledge [16]. "Big data" as a new information technology concept of the modern information society, science, industry and business, is associated with a wide range of diverse aspects of human activity. Information and communication technologies such as "Big data", cloud computing and widespread connection to telecommunication networks form a modern information and technology platform that automates the various processes of selection, storage, data processing and visualizing. Today there are several approaches to the definition of the term "Big data" and the presentation of its characteristics, in [17] the following definition is given: "the amount of data that is beyond the capacity of existing technologies to store, manage and efficiently process them" and in [16] another definition is given: "Big data represent a set of methods and technologies that require new forms of integration to uncover hidden knowledge in large data sets different in structure, complexity and characteristics".

In the paper [18], the updated concept of "Big data" is considered as "5V", which adds another parameter of veracity, which characterizes the possibility of classifying and processing data collections in the context of a particular process or model. In [19], the extended concept of "Big data" is proposed as "6V", which additionally contains such a characteristic as visualization of heterogeneous data in hybrid media spaces, and the concept "6V" presented in [20] includes another characteristic, which according to the authors' opinion should be variability, which characterizes changes in the speed of data flow, because it can fluctuate with unpredictable peaks and descents, which are caused by events in the real environment and are difficult to process with the help of limited computing resources. It is noted that in such situations, additional investment in computing resources to calculate peak levels of data collection may be ineffective due to insufficient use of such additional resources during the remaining time periods. The expansion of characteristics set of "Big data" concept to "7V" is presented in [21] and includes such characteristics as volume, velocity, veracity, variety, validity, value, visibility or visualization, virtuality, variability, or volatility, and in the paper [22] the concept of "Big data" is presented which is characterized as "7V" and contains the following features: volume, speed, velocity, veracity, variety, valence, variability, and value. According to the results of the analysis, the authors formed a list of characteristics of " Big data" which is presented in the format "10V" and contains the following attributes:

1. Volume – the total tuple of data types obtained from different types of sources, the list of which continues to expand. A significant advantage of the processes of "Big data" collection is the implementation on their basis of the procedures of search and extraction of hidden information and the formation of appropriate models based on the results of analytical data processing.
2. Velocity – characterizes the exponential growth of the rate of data accumulation, which in turn causes dynamic changes in collections and data content due to the addition of new collections, the use of archival or historical sets, connection of new streams of data from different sources and addition of collections of extracted sets of benefits data and knowledge, which in turn leads to high complexity and the cost of their processing [23].
3. Variety – characterizes the types of data collected by sensors, smartphones, tablets and PCs that contain audio files and video images, text, geospatial data, and numeric rows in structured or unstructured formats.
4. Validity refers to the relevance of data in the context of the method used for analytical work or process. An example can be outdated data in e-commerce systems that are not expedient to use due to their irrelevance.
5. Value characterizes the processes of detecting hidden values and knowledge in fast-moving multitype sets of "Big data" and involves reducing cost of operating expenses. The processes of processing "Big data" include identifying useful knowledge, calculating the return on investment, and determining the levels of relevance. "Big data" analytical processes can be implemented using a wide range of methods, tools and techniques, including traditional SQL queries, machine learning, data visualization, statistical and optimization methods, decision support procedures.
6. Veracity is a qualitative indicator of data collections that characterizes accuracy, quality, reliability, uncertainty, completeness of data collections and their suitability for analytical processes.
7. Visibility or visualization – characterizes the opportunities for effective presentation of important information in the "Big data" and implementing a comprehensive data analysis based on methods and algorithms for visual analytics in real time. [24]
8. Virtual – characterizes data management processes that are virtual.
9. Variability or volatility is a qualitative indicator of data collections that characterizes changes in the rate of obtaining qualitative and contextual data characteristics [23].
10. Valence – characterizes the possibility of representing "Big data" links in the form of graphs, that generates the need to use complex algorithms for analytical data processing [23].

Having determined the attributes of "big data", it is advisable to consider the methods of analytical processing of Big data in the projects of the "smart city" class.

METHODS FOR ANALYTICAL PROCESSING OF BIG DATA

The paper [25] presents the classification of software-algorithmic tools and means of "Big data" processing according to the purpose:

- Processing or computing: Hadoop, Nvidia CUDA or Twitter storm;
- Storage: Titan or HDFS;
- Analysis: MLPACK or mahout.

Tools for "Big data" processing can be classified according to the volumes of information sets:

- Whiteboard, R, MATLAB, octave – from kilobytes to several megabytes.
- Numpy, Scipy, Weka, Blas – from megabytes to several gigabytes.
- Hive, Mahout, Harna, Giraph – from gigabytes to terabytes.

The methods of analytical processing of "Big data" are classified as follows:

- Machine learning – involves the choice of analytical procedures for constructing models for assessing the effectiveness of the result. Traditionally, machine learning methods are divided into two categories: methods based on logical representations and statistical methods. The most common methods of machine learning that are used for "Big data". In particular, there are: a classification – to predict categories of input data, regression – to predict numeric values, clusterization – to organize such entities in a group, associative analysis – to search for connections between sets of entities, graph analysis – to identify links among objects using the structure of graphs, decision trees – to predict the statistical behavior of models for objective changes by analyzing decision-making rules formed on the basis of data functions. In analytical software tools, methods of managed and unmanaged machine learning are used. Managed machine learning includes in particular methods: regression (including linear regression), generalized linear model, ensemble methods, decision trees, neural networks, and classification algorithms such as vector constraint support, Bayesian analysis and the search for the closest neighbor. Unmanaged machine learning uses clustering techniques that include k-means clustering, kmedoids, fuzzy c-means, hierarchies, Gaussian mixtures, neural networks and hidden Markov models. In this case, the most varied software applications operating in real time are used.
- Statistical analysis – based on the use of tools and methods for collecting, analyzing and visualizing the results of processing large volumes of data collections, includes a variety of statistical methods and algorithms, used for cluster analysis, data mining and predictive modeling.
- Mining of "Big data" is a generally more complicated process in comparison with traditional methods of data mining, such as the detection and extraction of templates. In this case, statistical methods, methods of machine learning and pattern recognition are actively used [26], in particular methods of multiple regression of clustering algorithms, association analysis, and decision trees.

Thus, we have shown the system-technological connections of these classes of information technologies and analyzed the possibilities of their use in the context of implementation of the information technology project "Ternopil Smart city", which is carried out jointly by the scientists of Ternopil Ivan Puluj National Technical University and Lviv Polytechnic National University.

CONCLUSIONS

The paper presents the features of use of "fog" computing to build information technology platforms of a smart city using IoT devices. "Fog computing" is essentially a virtualized information technology platform that allows us to move part of the computing processes from the cloud information infrastructure of the"Smart city" to network resources close to the lower (sensory) level, to release computing power for analytical processing of urban data collections. Also, the properties and parametrization of the new information technology concept "Big data" are analyzed. In particular, there are ten attributes of "Big data" that can be identified as a separate class of information technology. The classification of software and algorithmic tools, means of processing "Big data" and methods of their analytical processing in accordance with the purpose and volume of information sets is presented. In the process of further research, it is necessary to design the architecture and structure of the information technology platform of "Smart city" using "Fog computing" and develop its software and algorithmic implementation with the possibility of analytical processing of the collected "Big data".

REFERENCES

1. **Gantz J., Reinsel D. 2012.** The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Analyze the future 2007, 1–16.
2. **Habak K. 2017.** 7 Elastic Mobile Device Clouds: Leveraging Mobile Devices to Provide Cloud Computing Services at the Edge, Fog for 5G and IoT.
3. **Habak K. 2017.** 7 Elastic Mobile Device Clouds: Leveraging Mobile Devices to Provide Cloud Computing Services at the Edge", Fog for 5G and IoT.
4. **Chiang M. 2017.** Clarifying Fog Computing and Networking: 10 Questions and Answers", IEEE Commun. Mag, Vol. 55. No. 4, 18–20.
5. **Zikopoulos P. 2013.** Harness the power of big data: The IBM big data platform. New York, NY: McGraw-Hill
6. Berman J. J. **2013.** Principles of big data: preparing, sharing, and analyzing complex information. Newnes.
7. Gartner IT Glossary > Big Data. Available online at: https://www.gartner.com/it-glossary/big-data.
8. **Gantz J., Reinsel D. 2011.** Extracting value from chaos.IDC iview 1142.2011, 1–12.

9. **Aazam Mohammad, Sherali Zeadally, and Khaled A. Harras. 2018.** Fog Computing Architecture, Evaluation, and Future Research Directions. IEEE Communications Magazine 56.5, 46–52.
10. **Aazam, Mohammad, Sherali Zeadally, Khaled A. Harras. 2018.** Fog Computing Architecture, Evaluation, and Future Research Directions. IEEE Communications Magazine 56.5, 46–52.
11. **Aazam, Mohammad, Sherali Zeadally, Khaled A. Harras. 2018.** Fog Computing Architecture, Evaluation, and Future Research Directions. IEEE Communications Magazine 56.5, 46–52.
12. **Manzalini A., Crespi N., 2016.** An Edge Operating System Enabling Anything-as-a-Service, IEEE Commun. Mag., Vol. 54. No. 3, 62–67.
13. **Habak K. 2017.** 7 Elastic Mobile Device Clouds: Leveraging Mobile Devices to Provide Cloud Computing Services at the Edge, Fog for 5G and IoT.
14. **Chiang M. 2017.** Clarifying Fog Computing and Networking: 10 Questions and Answers, IEEE Commun. Mag., Vol. 55. No. 4, 18–20.
15. **Hong K. 2013.** Mobile Fog: A Programming Model for Large-Scale Applications on the Internet of Things. Proc. 2nd ACM SIGCOMM Wksp. Mobile Cloud Computing, Aug. 2013, 15–20.
16. **Hashem Ibrahim Abaker Targio. 2015.** The rise of “big data” on cloud computing: Review and open research issues. Information Systems 47, 98–115.
17. **Manyika J. 2011.** Big data: The next frontier for innovation, competition, and productivity.
18. **Demchenko Yu., Cees De Laat, Membrey P. 2014.** Defining architecture components of the Big Data Ecosystem. Collaboration Technologies and Systems (CTS), 2014 International Conference on. IEEE, 2014.
19. **Ahn Jong Wook, Mi Sook Yi, Dong Bin Shin. 2013.** Study for spatial big data concept and system building. Journal of Korea Spatial Information Society 21.5, 43–51.
20. **Demchenko Yu., Gruengard E., Sander Klous. 2014.** Instructional model for building effective Big Data curricula for online and campus education. Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on. IEEE, 2014.
21. **Uddin Muhammad Fahim, Navarun Gupta. 2014.** Seven V's of Big Data understanding Big Data to extract value. American Society for Engineering Education (ASEE Zone 1), 2014 Zone 1 Conference of the. IEEE, 2014.
22. **Saggi Mandeep Kaur, Sushma Jain. 2018.** A survey towards an integration of big data analytics to big insights for value-creation. Information Processing & Management.
23. **Sivarajah Uthayasankar. (2017).** Critical analysis of Big Data challenges and analytical methods. Journal of Business Research 70, 263–286.
24. **Keim D., Huamin Qu, Kwan-Liu Ma. 2013.** Big-data visualization. IEEE Computer Graphics and Applications 33.4, 20–21.
25. **Tsai Chun-Wie. 2015.** Big data analytics: a survey. Journal of Big Data 2.1, p. 21.
26. **Chen CL Philip, Chun-Yang Zhang. 2014.** Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences 275, 314–347.