EDITOR: Schahram Dustdar, dustdar@dsg.tuwien.ac.at

DEPARTMENT: INTERNET OF THINGS, PEOPLE, AND PROCESSES

Elastic Data Analytics for the Cloud-to-Things Continuum

Sergio Laso, Global Process and Product Improvement, 10160 , Cáceres, Spain
Javier Berrocal, University of Extremadura, 06006 Badajoz, Spain
Pablo Fernandez and José María García, Universidad de Sevilla, 41004 Sevilla, Spain
Jose Garcia-Alonso and Juan M. Murillo, University of Extremadura, 06006 Badajoz, Spain
Antonio Ruiz-Cortés, Universidad de Sevilla, 41004 Sevilla, Spain
Schahram Dustdar, TU Wien, 1040 Vienna, Austria

The massive deployment of Internet-connected devices has led to an increase in the collection of data that are then used by companies to improve their decision-making processes. This growing trend demands more and more cloud and communications infrastructure. The limited resources, the need for sharing them, and the fact that many consumers are interested in the same data, call for an efficient management of the available resources. The cloud-to-things continuum can be used to execute different analytics closer to the data source so that infrastructure consumption and data circulation can be optimized. In this article, different dimensions for achieving elastic analytics and a framework for dynamically modifying their behavior are proposed.

Over the last few years, there has been a massive deployment of Internet of Things (IoT) devices, from heart rate sensors to location monitoring devices. This deployment has been especially driven by IoT applications that facilitate everyday tasks for every individual and organization. These tasks cover a wide range of use cases and applications, from managing an individual's personal health to understanding the flow of people and movement patterns in a city. This massive deployment of IoT devices has also led to the creation of networks of smart devices. This growing use of IoT devices is putting stress on current infrastructures in different dimensions.

On the one hand, it is putting stress on the amount of devices deployed. As IoT applications become more complex, they require combining devices from different domains for offering more useful functionalities. So far, new IoT applications and functionalities require the deployment of new devices. However, in this growing trend, the deployment of new IoT devices for sensing similar data is becoming unfeasible and unsustainable. For instance, a city that monitors the movement patterns of its citizens may also want to combine the location information with the heart rate in order to know what kind of activity citizens tend to do in each area of the city and, thus, provide more useful services and better plan their investments. For that purpose, instead of asking citizens to wear new heart rate monitors, it would be more feasible to ask them to share the data obtained by the monitors they already wear for other apps. Thus, in coming years, we will see how IoT applications will share large networks of devices already deployed.

On the other hand, it is putting stress on the infrastructure for data circulation and processing. Data, especially those coming from IoT devices, have become a strategic asset because of their availability to continuously monitor the environment without human intervention. IoT device networks are enablers of a new data economy in which more and more data are being exchanged not only within but also amongst

1089-7801 © 2022 IEEE
Digital Object Identifier 10.1109/MIC.2021.3138153
Date of current version 23 December 2022.

companies.[1] This fact boosts the massive deployment of IoT device networks. However, it also entails an increase in the circulation of information and the need for its processing and, consequently, a management model to support the optimal network usage. Such requirements will be even harder when different information systems require data from the same devices, for different needs and with different qualities. For example, a system to monitor an individual's heart rate during his/her activities may require very high information freshness; however, for a municipality to plan its services, such information does not have to be very fresh, but it has to be obtained from as many devices as possible.

In order to reduce the network overhead and improve the quality of service (QoS), IoT applications already attempt to take advantage of the cloud-to-things continuum to deploy services closer to information providers and consumers. However, such applications are still isolated systems that do not favor the sharing of data or analytics streaming. Thus, a sustainable data economy in this market[2] poses many challenges, among which we can count the following.

- First, the shared use of IoT devices so that data they are sensing can be used by different applications minimizing the drain on resources.
- Second, the optimization of the use of the cloud-to-things continuum infrastructure by enabling optimized deployment so IoT applications interested in the same data streaming and analytics can be deployed as close to the data as possible and together in the same nodes of this continuum.
- Third, elimination of duplicate streams by enabling applications interested in the same data or analytical streams to use the same datum.

In this article, we address these challenges by using elastic IoT data analytics whose behavior can be dynamically modified according to the quality requirements defined by each IoT system and the available resources of the cloud-to-thing infrastructure. For defining and deploying these analytics, we also provide a framework with an orchestrator managing the elasticity. This orchestrator evaluates the state of the infrastructure, the other analytics in progress, and the QoS required by each analytic. As a result, it reconfigures all the analytics to maximize the quality provided by each of them without exhausting the infrastructure nor the IoT devices.

In the following, we first discuss the different dimensions that should be taken into account for building elastic data analytics. Then, we present a framework for achieving elastic data analytics considering some of the defined dimensions. Finally, we present some conclusions and discussion about the elasticity required by IoT applications.

ENABLING ELASTIC DATA ANALYTICS

Data analytics allow companies to process large volumes of information in order to: find trends, eliminate unnecessary information, aggregate information, etc. In addition, they can enable efficient information management, as they may reduce the overhead of the transferred data. However, they may also require data storage and processing capabilities from the cloud-to-things continuum nodes. These analytics must be defined by each data consumer, since the results must be adapted to their requirements. Nevertheless, the circulation of information should be controlled to increase the efficiency in the use of both the system and the data.

Let us define a smart city scenario to better show the need of elastic data analytics. In order to maintain the simplicity, this case study is focused only on the citizens' location. Different institutions can analyze this information to identify their movement patterns, but with different goals. For instance, a healthcare system can use the real-time movements of the citizens to identify crowded areas and reduce the risk of infection for COVID-19. Instead, the city council may require long-term movement patterns to study future infrastructure investments (accessibility, bus lines, etc.).

The execution of these analytics impacts on several highly related dimensions, as Figure 1(a) shows, that must be controlled in a sustainable computing and information environment: data quality, resource consumption and cost.

First, one of the most important aspects for data consumers is the quality of the information. This quality must be appropriate for the specific functionalities they want to provide. While there are different properties to measure data quality, for information coming from IoT devices two properties are particularly relevant.[3]

1) *Accuracy:* It is the level to which data represent the real-world scenario. In this sense, in a network of IoT devices, the higher the number of devices involved in the data analytic the better they will represent the real world.
2) *Freshness* or the timeliness of the data: For some applications, the data have no value if it is

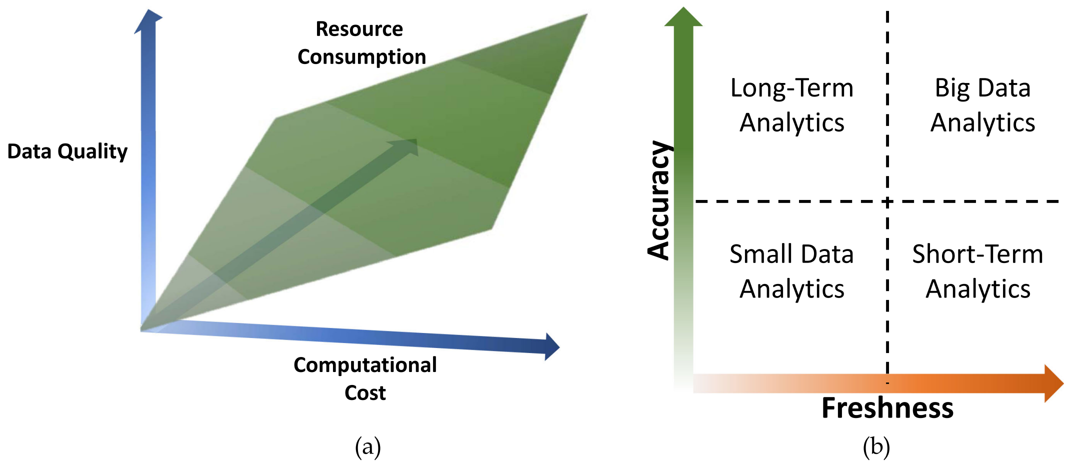


FIGURE 1. (a) Dimensions impacting the analytics. (b) Analytics depending on the data quality.

not available at the right moment. Freshness indicates the frequency at which IoT devices have to provide information for the analytic to provide value to the consumer.

Combining both dimensions, as Figure 1(b) shows, different types of analytics can be executed. When the accuracy and the freshness are low, small data analytics are executed in order to have a limited quantity of highly granular data that usually provide valuable information for the system. As the freshness and the accuracy increase, bigger data analytics are executed focusing on processing large volumes of information for business decisions. Instead, if only the accuracy increases, long-term data analytics are executed for predictive decision-making processes. In addition, if mainly the freshness is important for achieving an acceptable data quality, short-term data analytics are usually executed for developing reactive processes. For instance, for our case study, the healthcare system would require short-term analytics while the city council would require long-term analytics.

Second, this type of analytics has a direct impact on the resource consumption of the nodes in the cloud-to-things continuum. A greater accuracy means that more IoT devices will be involved in sending information, consuming their resources, and increasing the number of information flows. Edge and fog nodes can be used to distributedly process such analytics, reducing and aggregating the information flows. However, the greater the dispersion, the greater the number of nodes involved. Therefore, a high accuracy usually leads to a greater distribution of the resource consumption over the entire network. For our case study, the city council may require to use different nodes to cover the whole city, storing and aggregating the movement patterns whilst the healthcare system would only require a sample of some areas.

Finally, in addition to the data cost (which may depend on each data producer), its processing also entails some infrastructure costs associated with the use of edge and fog nodes.[4] Thus, the greater the freshness and frequency at which the information must be processed, the greater the need for processing this information in the fog or edge nodes, and the higher the infrastructure cost of the analytics. This cost may be limited to a small number of fog or edge nodes if the accuracy is low or to a larger one as the accuracy increases. For instance, for the healthcare analytics only the nodes involved in crowded areas would incur a higher cost. Furthermore, in a leveraged scenario with an open market of sensors,[5] this cost could also be dynamic depending on the fluctuations of the market.

In environments where multiple IoT applications are consuming similar information and using the same computing resources, data analytics should be able to have an elastic behavior.[6] They should be able to increase or decrease the provided quality depending on the available resources (and within the consumer's requirements). In addition, a framework would be needed to efficiently manage the distribution of all the requested analytics on the cloud-to-things continuum. For each analytic, this framework should identify in which fog or edge nodes it should be deployed to meet the accuracy and freshness requirements, without overloading the infrastructure. Likewise, when a new analytic is requested, the framework should use the defined elasticity to reconfigure the already existing analytics in order to achieve an efficient use of the

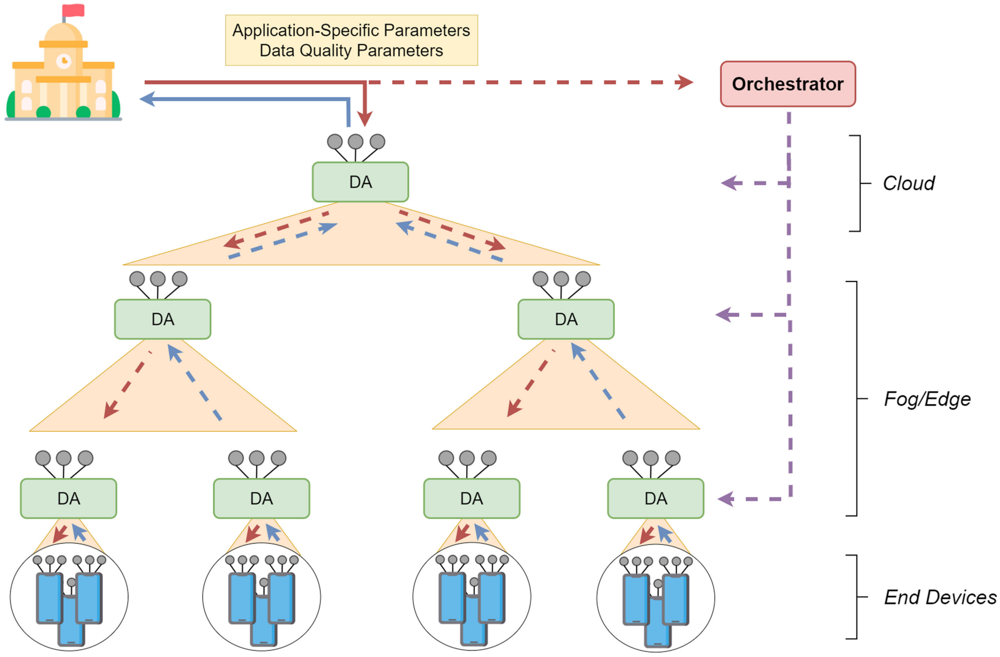


FIGURE 2. Architecture of the elastic data analytic framework.

computing resources, while maximizing the provided quality. Moreover, this distribution should also consider if several analytics should be deployed together, because they can share the same sensed data and, thus, increase the general efficiency of the system.

REALIZATION AND APPLICATION OF THE FRAMEWORK

To achieve elastic data analytics, we propose a framework orchestrating the analytics required by different data consumers and leveraging the cloud-to-things continuum to distribute them depending on the available resources and the required quality. The source code of the framework[a] for the presented case study and a video[b] showing the achieved elasticity are publicly available.

Figure 2 shows an example of the infrastructure that can be used for our smart city case study. It shows the different layers of the cloud-to-things continuum. Note that we limited the number of layers to improve readability. Data consumers (healthcare system and city council, for our case study) can request different analytics through the entry point to the infrastructure, the cloud. The deepest layer is composed by the end devices sensing and sending information to the cloud. This information goes through the fog and edge nodes on its way to the cloud where it is provided to data consumers.

Data analytics can be requested with different parameters. First, the application-specific parameters (for instance, the area to monitor in the smart city) and, second, the data quality parameters (accuracy and freshness). For the current implementation, the accuracy and the freshness can take different values from a range (low, medium, and high). The accuracy relates to the ratio of end devices involved, and the freshness to the frequency at which the information is obtained.

Figure 2 also defines the most important components of the proposed framework. First, cloud, fog, and edge nodes are composed by a data aggregator (DA) component. This component processes the deployed analytics. To that end, it requests the required information to the lower nodes depending on the data quality (mainly the freshness) required by the most demanding analytic, caches the obtained data to be reused by the other analytics, and processes it. In addition, the obtained results are also cached and always available, waiting for the higher levels to request them.

In addition, an elasticity orchestrator is proposed to balance the load of the whole infrastructure, modify the behavior of the elastic analytics depending on the previously defined dimensions, and to provide the best QoS to all parties. When a new analytic is required, it is analyzed by the orchestrator to evaluate the desired quality

[a][Online]. Available: https://doi.org/10.5281/zenodo.5793214
[b][Online]. Available: https://doi.org/10.5281/zenodo.5793189

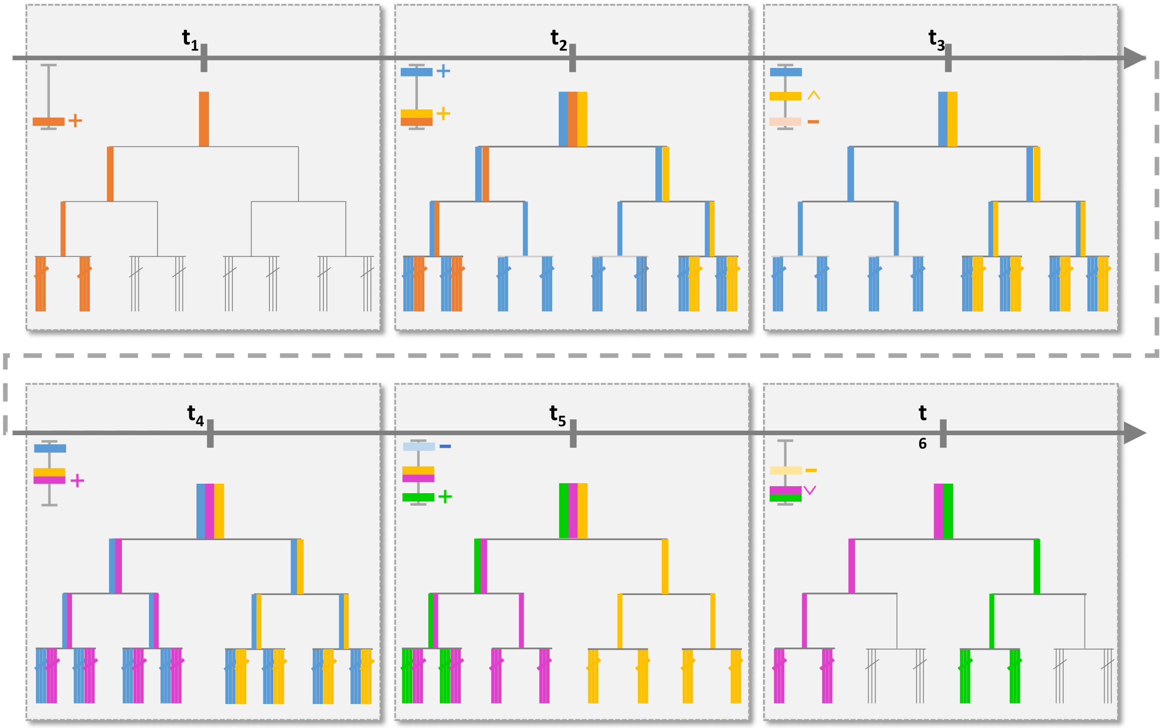
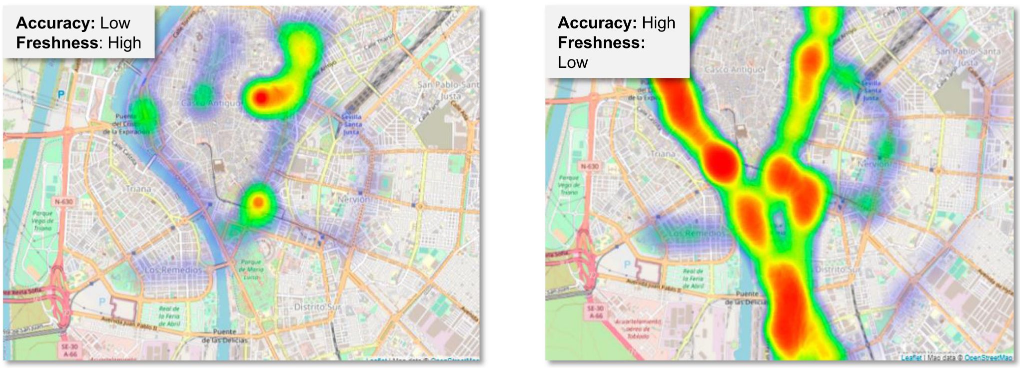


FIGURE 3. Elasticity dynamics.

(accuracy and freshness), the analytics already deployed in the architecture, and the workload of the different nodes. If there are enough resources, it directly deploys it providing the highest possible quality. If there are not enough resources, it checks if the analytics already deployed can be reconfigured to make room for the new one. This reconfiguration can be provided by redeploying existing analytics on other nodes to have a more efficient distribution or, if possible due to the quality defined by consumers, reducing the accuracy and freshness of some analytics to free up some resources.

Finally, all the DAs have the same communication interface (API). Thus, the infrastructure can be easily increased or reduced without heavily impacting the orchestrator. This also facilitates the work of developers and operators, avoiding having to create *ad hoc* solutions.

Figure 3 shows a working example of the provided elasticity. The example consists of a timeline with six



Analytic 1: Healthcare system, identification of crowded areas

Analytic 2: City council, study of movement patterns

FIGURE 4. Analytics obtained for the smart city case study.

DATA MANAGEMENT IN THE CLOUD-TO-THINGS CONTINUUM

Cloud computing has not only been a revolution at the business level to reduce IT costs. It has also fostered the massive deployment of new Internet-connected devices, with reduced computing resources but with huge capabilities for sensing and interacting with the environment. This has led to a steady increase in the data flows between these connected devices and the cloud. Paradigms such as fog and edge computing have brought the cloud environments, and hence the information processing, closer to the data sources. In fog computing, the processing is done on servers residing between the cloud and end devices, while edge computing focuses on processing the information at the edge of the network (gateways and end devices). By processing the information closer to the data source, the infrastructure overhead and the QoS can be improved.[1]

These paradigms allow the generation of distributed clouds making use of the cloud-to-thing continuum for the deployment of services throughout the entire infrastructure.[2] These services, some of them focused on data processing, can dynamically migrate from one layer to another depending on the volume of information to be processed, the workload of the nodes, or the QoS desired. Different technologies even orchestrate these distributed services to execute complex workflows.[3]

The incorporation of intelligence to the cloud-to-thing continuum is allowing researchers to take a further step in the processing of these huge data flows in order to get predictions and automate the decision-making process. However, fog and edge nodes have limited computing resources, hindering the execution of common artificial intelligence techniques. New techniques are being defined for reducing the rigidity of these mechanisms, allowing their distribution along the cloud-to-thing infrastructure.[4]

Finally, the cloud-to-thing continuum should not only manage the distribution of resources for a single IoT application but also support the integration of various applications, even from different domains. However, the different data semantics, requirements, and qualities of the different consumers complicate the interoperability and turn the collaboration across multiple vertical systems into a challenging task. New architectures are being defined to manage the infrastructure resources and share services between different systems.[5] Nevertheless, an elastic management of the information flows in the cloud-to-thing continuum is needed. This elasticity must consider the limited infrastructure resources, the interactions between applications, the data that can be shared between them, and the quality required.

REFERENCES

1. T. Lynn, J. G. Mooney, B. Lee, and P. T. Endo, *The Cloud-to-Thing Continuum: Opportunities and Challenges in Cloud, Fog, and Edge Computing*. Berlin, Germany: Springer, 2020, doi: 10.1007/978-3-030-41110-7.
2. J. L. Herrera, J. Galán-Jiménez, J. Berrocal, and J. M. Murillo, "Optimizing the response time in SDN-fog environments for time-strict IoT applications," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 17172–17185, Dec. 2021, doi: 10.1109/JIOT.2021.3077992.
3. M. Villari, M. Fazio, S. Dustdar, O. Rana, D. N. Jha, and R. Ranjan, "Osmosis: The osmotic computing platform for microelements in the cloud, edge, and Internet of Things," *Computer*, vol. 52, no. 8, pp. 14–26, 2019, doi: 10.1109/MC.2018.2888767.
4. D. R. Torres, C. Martín, B. Rubio, M. Díaz, "An open source framework based on Kafka-ML for DDNN inference over the cloud-to-things continuum," *J. Syst. Archit.*, vol. 118, 2021, Art. no. 102214, doi: 10.1016/j.sysarc.2021.102214.
5. N. Chen, Y. Yang, J. Li, and T. Zhang, "A fog-based service enablement architecture for cross-domain IoT applications," in *Proc. IEEE Fog World Congr.*, 2017, pp. 1–6, doi: 10.1109/FWC.2017.8368533.

instants (*t*). At each instant, an action is triggered that has an impact on the infrastructure. At the top left of each instant, a small legend appears with a vertical bar. This legend indicates for each analytic (each colored horizontal bar) if it is inserted/removed from the system (+/−) and, depending on its height in the vertical bar, the accuracy that it will provide. Note that we only represent the accuracy dimension to improve readability. At *t1*, the orange analytic is requested, requiring low accuracy. The orchestrator, therefore, assigns it to only a part of the infrastructure. At *t2*, two new analytics (yellow and blue) are inserted, with low and high accuracy, respectively; the orchestrator deploys the blue analytic throughout the whole infrastructure while the yellow one is deployed only in the less loaded part. At *t3,* the orange analytics ends and, because there are enough resources, the orchestrator upgrades the yellow one to a medium accuracy. At *t4,* the purple analytic is inserted and the orchestrator again assigns it the lowest loaded part of the infrastructure. At *t5*, the blue analytic ends and a new one with a low accuracy is deployed. Finally, at *t6*, the yellow analytics ends and the rest are reorganized to better distribute the resource consumption.

Finally, to show how the elasticity may affect the results obtained, Figure 4 shows two different analytics for the smart city case study. For each analytic, the data quality parameters (freshness and accuracy) are detailed. The results are the citizens' locations, which are shown with heatmaps for an easier decision-making process. The first analytic is focused on quickly identifying crowded areas for the healthcare system. Therefore, it has been requested with low accuracy and high freshness. Consequently, the orchestrator has only ordered the execution of the analytics to a part of the infrastructure and fewer citizens are involved but with real-time data. On the other hand, the second analytic would be interesting for the city council to study future investments. In this case, it has been requested with high accuracy and low freshness. Therefore, the orchestrator assigns the entire infrastructure to execute them and, therefore, gets the information from all the available citizens in the area.

CONCLUSION

The deployment of a swarm of sensors and Internet-connected devices is giving rise to a new data-driven economy. We envision an environment in which both the data to be consumed and the existing resources must be efficiently managed. In this article, we propose elastic data analytics, and we define the different dimensions affected by them. We also outline a framework to manage the dynamicity of these analytics in real time depending on the context.

Nevertheless, there are still some open challenges to be addressed to fully implement elastic data analytics. First, different policies should be generated for changing the management of elasticity depending also on the cost. For instance, trying to maximize the available computing resources, or minimize the costs to foster a sustainable economy. Second, more dimensions should be analyzed to identify how the elasticity affects them, such as the cost of the data. Currently, we work on including artificial intelligence to define self-managed analytics.

ACKNOWLEDGMENTS

This work was supported in part by the DIN2020-011586 Grant, funded by MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU/PRTR," in part by the Projects RTI2018-094591-B-I00, RTI2018-101204-B-C21 (MCIU/AEI/FEDER, UE) and 0499_4IE_PLUS_4_E (Interreg V-A España-Portugal 2014-2020), in part by the Department of Economy, Science and Digital Agenda of the Government of Extremadura under Grants GR18112, IB18030 in part by Junta de Andalucía under projects APOLO (US-1264651) and EKIPMENT-PLUS (P18-FR-2895), in part by the European Regional Development Fund, and in part by the RCIS Network under Grant TIN2016-81978-REDT.

REFERENCES

1. A. Opher, A. Chou, A. Onda, and K. Sounderrajan, *The Rise of the Data Economy: Driving Value Through Internet of Things Data Monetization*. New York, NY, USA: IBM Corporation: Somers, 2016.
2. F. Xhafa, B. Kilic, and P. Krause, "Evaluation of IoT stream processing at edge computing layer for semantic data enrichment," *Future Gener. Comput. Syst.*, vol. 105, pp. 730–736, 2020, doi: 10.1016/j.future.2019.12.031.
3. J. Medd, "Data quality fundamentals for data and analytics technical professionals," Gartner, Aug. 17, 2020. [Online]. Available: https://www.gartner.com/en/documents/3989182/data-quality-fundamentals-for-data-and-analytics-technic
4. A. Brogi, S. Forti, and A. Ibrahim, "Deploying fog applications: How much does it cost, by the way?," *Small*, vol. 1, no. 2, 2018, Art. no. 20, doi: 10.5220/0006676100680077.
5. J. M. García, P. Fernandez, A. Ruiz-Cortés, S. Dustdar, and M. Toro, "Edge and cloud pricing for the sharing economy," *IEEE Internet Comput.*, vol. 21, no. 2, pp. 78–84, Mar./Apr. 2017, doi: 10.1109/MIC.2017.24.
6. S. Dustdar, P. Fernandez, J. M. García, and A. Ruiz-Cortés, "Elastic smart contracts in blockchains," *IEEE CAA J. Autom. Sinica*, vol. 8, no. 12, pp. 1901–1912, Dec. 2021, doi: 10.1109/JAS.2021.1004222.

SERGIO LASO is currently with Global Process and Product Improvement, Cáceres, Spain, where he is working toward the industrial Ph.D. degree in computer science. Contact him at slasom@unex.es.

JAVIER BERROCAL is currently an associate professor with the Department of Informatics and Telematics System Engineering, University of Extremadura, Badajoz, Spain. Contact him at jberolm@unex.es.

PABLO FERNANDEZ is currently an associate professor with the Applied Software Engineering Group, Smart Computer Systems Research and Engineering Lab, Research Institute of Computer Engineering, University of Sevilla, Seville, Spain. Contact him at pablofm@us.es.

JOSÉ MARÍA GARCÍA is currently an associate professor with the Applied Software Engineering Group, Smart Computer Systems Research and Engineering Lab, Research Institute of Computer Engineering, University of Sevilla, Seville, Spain. Contact him at josemgarcia@us.es.

JOSE GARCIA-ALONSO is currently an associate professor with the Department of Informatics and Telematics System Engineering, University of Extremadura, Badajoz, Spain. Contact him at jgaralo@unex.es.

JUAN M. MURILLO is currently a full professor of Software Engineering with the University of Extremadura, Badajoz, Spain. Contact him at juanmamu@unex.es.

ANTONIO RUIZ-CORTÉS is currently a full professor and head of the Applied Software Engineering Group (ISA) with the Smart Computer Systems Research and Engineering Lab, Research Institute of Computer Engineering, University of Sevilla, Seville, Spain. Contact him at aruiz@us.es.

SCHAHRAM DUSTDAR is currently a full professor of computer science (informatics) with the Distributed Systems Group, TU Wien, Vienna, Austria. Contact him at dustdar@dsg.tuwien.ac.at.