SeDaSOMA: A Framework for Supporting Serendipitous, Data-As-A-Service-Oriented, Open Big Data Management and Analytics

Alfredo, Ac, Cuzzocrea*
iDEA Lab, University of Calabria, Rende, Italy & LORIA, Nancy, France

Paolo, Pc, Ciancarini
DISI Department, University of Bologna, Bologna, Italy

ABSTRACT

This paper describes the anatomy of SeDaSOMA, a reference framework for supporting *serendipitous, data-as-a-service-oriented, open big data management and analytics*. The proposed framework aims at supporting *advanced big data management and analytics* by relying on innovative research findings and next-generation big data tools. The paper also depicts some *Cloud-aware big data vertical applications* of SeDaSOMA in specific scenarios that are currently of great interest.

CCS CONCEPTS

• **Information systems** → Data management systems; Database management system engines; Data management systems; Information integration; Data management systems; Database management system engines; Database query processing.

KEYWORDS

Big Data, Big Data Management, Big Data Analytics, Serendipitous Big Data Methodologies, Big Data As a Service, Open Big Data

ACM Reference Format:

Alfredo, Ac, Cuzzocrea and Paolo, Pc, Ciancarini. 2021. SeDaSOMA: A Framework for Supporting Serendipitous, Data-As-A-Service-Oriented, Open Big Data Management and Analytics. In *2021 5th International Conference on Cloud and Big Data Computing (ICCBDC) (ICCBDC 2021), August 13–15, 2021, Liverpool, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3481646.3481647

1 INTRODUCTION

Big data management and analytics (e.g., [1-6]) are now consolidated research topics that combine together past experiences maturated in the contexts of database systems, data warehousing, data mining, and machine learning. Nowadays, big data management and analytics arise in a plethora di application scenarios, ranging from smart cities (e.g., [7]) to social networks (e.g., [8]), from sensor networks (e.g., [9]) to intelligent transportation systems (e.g., [10]), and so forth.

Big data expose several interesting research challenges, mostly deriving from the well-known *3V nature* [11], which identifies data repositories that are characterized by high volume, high velocity and high variety. Among these, a great deal of interest has been devoted to the issue of managing and supporting analytics over big data effectively and efficiently.

Indeed, from early research experiences, e.g. *OLAP data cube processing* (e.g., [12-17]) for big data management, and *OLAP mining* (e.g., [18-20]) for big data analytics, it follows that achievements and results cannot be applied "as they are" to the big data settings, but novel challenges arise and, therefore, innovative models, techniques and algorithms are necessary to this end (e.g., [21-24]). This requirement is magnified by the tremendous need for *intelligent big data applications* arising in the modern information and communication society.

Following this exciting research trend, this paper introduces the anatomy of SeDaSOMA, a reference framework for supporting Se*rendipitous,* Da*ta-as-a-Service-oriented,* O*pen big data* M*anagement and* A*nalytics.* The proposed framework aims at supporting *advanced big data management and analytics* by relying on innovative research findings and next-generation big data tools.

SeDaSOMA realizes a composite software architecture devoted to the serendipitous, Data-as-a-Service-oriented, open management of big data, and to predictive analytics over big data. This general goal pursues a collection of models, techniques and algorithms for acquiring, representing, managing and making-secure big data via modern *Cloud-based DaaS* (Data-as-a-Service) paradigms. The main goal consists in defining a modern data marketplace applicative setting for achieving an open big data environment that ultimately supports the definition and the implementation of scalable big data analytics, based on innovative (big) data query answering techniques, context-aware methods and (big) data preparation for analytics solutions. The main layers realized within the SeDaSOMA reference architecture represent the functional and procedural basis, typically oriented to data-intensive tasks, for a wide spectrum of *Cloud-aware big data vertical applications* ranging in the scope defined by the major areas of actual information society challenges drawn by the European data management guidelines [25], with particular reference to: (*i*) *social big data management for workplace safety and health*; (*ii*) *integration of (big) bank data and (big) customer data for supporting big data intelligence*; (*iii*) *big data advertisement on the Web for opportunity finding*. The latter ones are just case studies recognized as critical examples in modern big data settings.

The paper is organized as follows. In Section 2, we provide an analysis of related work that is relevant for our research. Section 3

*This research has been made in the context of the Excellence Chair in Computer Engineering – Big Data Management and Analytics at LORIA, Nancy, France.

ICCBDC 2021, August 13–15, 2021, Liverpool, United Kingdom

ACM ISBN 978-1-4503-9040-8/21/08…$15.00
https://doi.org/10.1145/3481646.3481647

focuses the attention on a detailed description of the SeDaSOMA framework and all its components. In Section 4, we provide a discussion on the research contributions of the SeDaSOMA proposal. Finally, in Section 5, we report conclusions and future work of our research.

2 RELATED WORK

As mentioned in Section 1, there is strong interest in the context of big data management and analytics context. Here, we review some relevant approaches.

[26] recognized that big data analytics and knowledge management were becoming a hot topic with the emerging techniques of cloud computing and big data computing model such as *MapReduce*. However, large-scale adoption of MapReduce applications on public clouds is hindered by the lack of trust on the participating virtual machines deployed on the public cloud. In the paper, authors extended the existing hybrid *Cloud MapReduce* architecture to *Multiple Public Clouds*. Based on such architecture, *IntegrityMR*, an integrity assurance framework for big data analytics and management applications was proposed. Authors explored the resulting integrity check techniques at two alternative software layers: the *MapReduce* task layer and the applications layer. They designed and implemented the system at both layers based on *Apache Hadoop MapReduce* and *Pig Latin*, and performed a series of experiments with popular big data analytics and management applications such as *Apache Mahout* and *Pig* on commercial public clouds (*Amazon EC2* and *Microsoft Azure*) and local cluster environments. The experimental result of the task layer approach shows high integrity (98% with a credit threshold of 5) with non-negligible performance overhead (18% to 82% extra running time compared to original *MapReduce*).

[27] describes the *Ophidia* project, a research effort addressing big data analytics requirements, issues, and challenges for *e*-science. Authors present here the *Ophidia* analytics framework, which is responsible for atomically processing, transforming and manipulating *array-based data*. This framework provides a common way to run on large clusters analytics tasks applied to big datasets. The paper highlights the design principles, algorithm, and most relevant implementation aspects of the Ophidia analytics framework. Some experimental results, related to a couple of data analytics operators in a real cluster environment, are also presented.

[28] argues that Cloud computing and big data analysis are gaining lots of interest across a range of applications including *disaster management*. These two technologies together provide the capability of real-time data analysis not only to detect emergencies in disaster areas, but also to rescue the affected people. The paper presents a framework that supports *emergency event detection and alert generation by analyzing the data stream*, which includes efficient data collection, data aggregation and alert dissemination. One of the goals for such a framework is to support *an end-to-end security architecture* to protect the data stream from unauthorized manipulation as well as leakage of sensitive information. The proposed system provides support for both data security punctuation and query security punctuation. The paper presents the proposed architecture with a specific focus on data stream security. It also briefly describes the implementation of security aspects of the architecture.

[29] focuses the attention on decision making in *natural disaster management*, which has its own challenges that need to be tackled. In times of disaster, government as a response organisation must conduct timely and accurate decisions to ensure rapid assistance and effective recovery for the victims involved can be conducted. The aim of the paper is to embark strategic decision making in government concerning to disaster management through *Big Data Analytics* (BDA) approach. BDA technology is integrated as a solution to manage, utilise, maximise, and expose insight of climate change data for dealing water related natural disaster. NAHRIM as a government agency responsible in conducting research on water and its environment proposed a BDA framework for natural disaster management using NAHRIM historical and simulated projected hydroclimate datasets. The objective of developing this framework is to assist the government in making decisions concerning disaster management by fully utilised NAHRIM datasets. The BDA framework that consists of three stages; *Data Acquisition*, *Data Computation*, and *Data Interpretation* and seven layers; *Data Source*, *Data Management*, *Analysis*, *Data Visualisation*, *Disaster Management*, and *Decision* is hoped to give impact in prevention, mitigation, preparation, adaptation, response and recovery of water related natural disasters.

[30] considers that *advanced digitalization* together with the rise of disruptive Internet technologies are key enablers of a fundamental paradigm shift observed in *industrial production*. This is known as the fourth industrial revolution (*Industry 4.0*) which proposes the integration of the new generation of ICT solutions for the monitoring, adaptation, simulation, and optimisation of factories. With the democratization of sensors and actuators, factories and machine tools can now be sensorized and the data generated by these devices can be exploited, for instance, to optimise the utilization of the machines as well as their operation and maintenance. However, analyzing the vast amount of generated data is resource demanding both in terms of computing power and network bandwidth, thus requiring highly scalable solutions. The paper presents *a novel big data approach and analytics framework for the management and analysis of machine generated data in the cloud*. It brings together standard open source technologies and the exploitation of *elastic computing*, which, as a whole, can be adapted to and deployed on different cloud computing platforms. This enables reducing infrastructure costs, minimizing deployment difficulty, and providing on-demand access to a virtually infinite set of computing power, storage and network resources.

Finally, [31] recognizes that, with an increased usage of ICT technologies in the smart cities domain, the data generated has increased manifolds. This data is *heterogeneous* in nature, as it varies with respect to time and exhibits the properties of all essential *V*'s for big data. Therefore, to handle such an enormous amount of data, various big data processing techniques are required. To cope up with these issues, the paper presents *a tensor-based big data management technique* to reduce the dimensionality of data gathered from the *Internet-of-Energy* (IoE) environment in a smart city. The core data is extracted out of the gathered data by using tensor operations such as-matricization, vectorization and tensorization with the help of higher-order singular value decomposition. This core data is then stored on the cloud in the reduced form. After reducing the dimensionality of data, it is used for providing many services in smart cities; and its application to provide *Demand Response* (DR) services has been discussed in the paper. For this purpose, *Support Vector Machine* (SVM)-based classifier is used to classify the end-users (residential and commercial) into normal, overloaded and underloaded categories from the core data. Once such users are identified to take part in DR mechanism, utilities then generate commands to handle their DR in order to alter load requirements so that the overall load is optimized. Results obtained on *Open Energy Information* and *PJM* dataset clearly indicate the supremacy of the proposed tensor-based scheme over the traditional scheme for DR management.

3 SEDASOMA'S ANATOMY

Figure 1 depicts the SeDaSOMA layered framework architecture. The main challenge of the SeDaSOMA framework consists in supporting and experimenting an innovative vision for big data management and analytics, which breaks the traditional way of addressing these issues, and achieves novel paradigms, mechanisms, and schemes to this end. The main focus of the SeDaSOMA framework is represented by the issue of *supporting scalability and responsiveness of analytics over big data*, and testing so-defined achievements in the context of critical Cloud-aware big data vertical applications falling in major areas of actual information society challenges drawn by the European guidelines.

As shown in Figure 1, the SeDaSOMA architecture is typically multi-layered, and, in a nested manner, each layer comprises, in turn, several components. In particular, four layers are introduced:

- *Big Data Source Layer*, which is the layer where the sources of big data, which are inherently streaming, distributed and heterogeneous in nature, are located and managed;
- *Big Data Repository and Provisioning Layer*, which is the layer where big data are stored, processed and provisioned to the higher layers;
- *Big Data Analytics Layer*, which represents the layer where useful knowledge is extracted from the big data via complex and scalable analytics running over these huge amounts of data;
- *Big Data Application Layer*, which represents the layer including Cloud-aware big data vertical applications, which completely rely on the underlying layers, are located – these applications focus the attention on major areas of actual information society challenges drawn by European Union data management guidelines.

Each layer is devoted to a particular collection of big data management and analytics issues (which, however, are already investigated in current literature) with the specific goal of achieving innovative and critical knowledge and research advancements, along with authoritative implementations via ad-hoc Cloud-aware vertical applications over big data. In addition to this, the SeDaSOMA framework exposes an organization, i.e. the relationships among the different layers composing the architecture, which is innovative itself, i.e. it introduces a novel and sustainable vision of managing and analyzing big data that has not similarities in actual literature.

3.1 SeDaSOMA Components

Now, focus the attention on Figure 1. Each layer of the SeDaSOMA framework is organized in terms of several components, as follows.

The Big Data Source Layer contains the following components:

- *Physical Resource Management.* This component deals with physical aspects of resources that may produce data streams (e.g., sensors and datacenters) by deeply looking into innovative aspects such as noise-detection and cleaning, stream pipelining, resource/stream synchronization, and so forth. Advanced topics include the design of high-performance infrastructures for dealing with typical challenges of big data streams, such as the well-known 3V, i.e. *Volume*, *Variety*, *Velocity*.
- *Heterogeneous Distributed Big Data Stream Management.* This component is concerned with the big data (stream) management at the framework's input interface, under the big-data's meaning as data objects on top of the underlying physical resources (e.g., sensors and datacenters that may produce big data streams). Here, the main issues to be faced-off are the following ones: (*i*) coping with the streaming nature of big data, (*ii*) dealing with multi-rate arrivals and inter-arrivals, (*iii*) dealing with massive amounts of (streaming) big data, and so forth.
- *Big Data Crowdsourcing.* This component deals with the problem of collecting big data by means of the innovative crowdsourcing paradigm. Basically, the main goal here consists in devising models, techniques and algorithms for applying this novel paradigm to the big data collection/acquisition problem.

The Big Data Repository and Provisioning Layer contains the following components:

- *Big Data Storage.* This component represents a critical component for the entire SeDaSOMA framework, as it directly deals with big data representation and storage issues, being scalability of the major requirements. Several issues arise here: from advanced data structures for storing big data to big data indexing, from big data partitioning methods to innovative data management policies aiming at making big data storage solutions elastic, and so forth.
- *Big Data Warehousing.* Due to the complex nature of big data, warehousing methodologies are the most convenient solutions capable of providing a reliable data management layer for the SeDaSOMA framework. Combined with the innovative Cloud-based paradigm, these methodologies are captured and implemented by the Big Data Warehousing component, which, in turn, founds on the Big Data Storage component. In particular, the component will be implemented via modern *NoSQL* architectures (e.g., in distributed settings). In this component, main issues are: (*i*) multidimensional models for big data, which are necessary in order to support big data analytics; (*ii*) compressed representations of big multidimensional data; (*iii*) MapReduce-based warehousing of big data, and so forth.
- *Secure Big Data Processing.* A critical requirement of the SeDaSOMA framework consists in guaranteeing the security of big data stored in the Big Data Warehousing component,

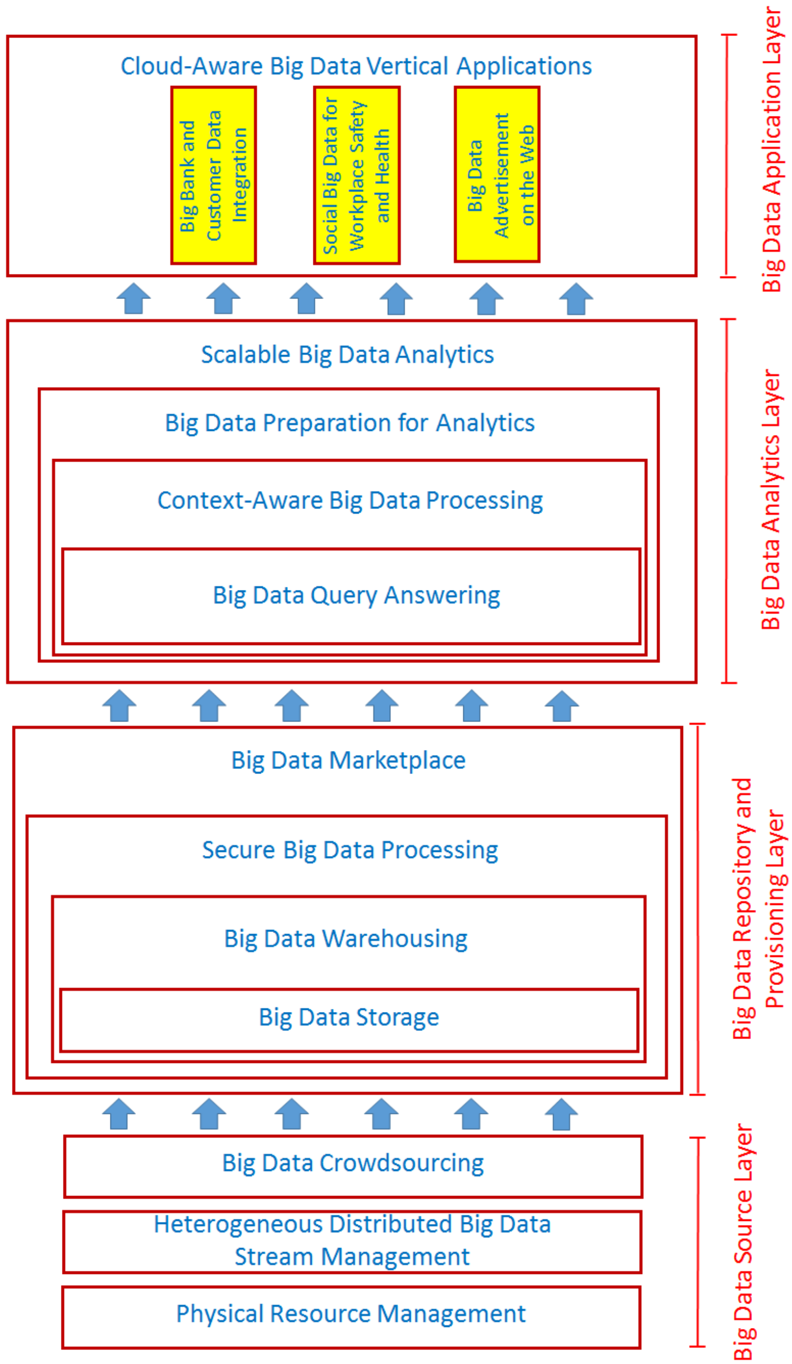

Figure 1: The SeDaSOMA Framework

due to the fact these (big) data are accessed and processed within an open environment represented by the Big Data Marketplace component. As a consequence, the SeDaSOMA framework must encompass ad-hoc solutions for making big data processing secure. This goal is realized by the Secure Big Data Processing component, which directly interfaces the Big Data Warehousing component. In this context, relevant issues are: (*i*) guaranteeing scalable secure access to big data; (*ii*) devising innovative encryption mechanisms for big data; (*iii*) defining data provenance methods for big data, and so forth.

- *Big Data Marketplace*. One of the innovative characteristics of the SeDaSOMA framework is represented by the fact that the core data layer provides and makes available big data to higher layers and consumer (Cloud-aware) applications via the novel data marketplace (DaaS) paradigm, rather than adopting a traditional big data fruition scheme. This paradigm copes with novel Cloud computing environments, and foresees an applicative setting where big data are made available to consumer applications via service-oriented primitives. Basically, the final goal of this component consists in making available big data stored in the Big Data Warehousing component to consumer applications according to the DaaS paradigm, in a secure manner (provided by the Secure Big Data Processing component). Several open issues arise here: from innovative service-oriented big data provisioning to models for big data marketplace, from scalability issues to Cloud-compliant heterogeneity issues, and so forth.

The Big Data Analytics Layer contains the following components:

- *Big Data Query Answering*. This component deals with the problem of devising effective and efficient algorithms for supporting query answering over big data, yet facing-off the scalability challenge. Classical query algorithms proposed and developed in the context of traditional data-intensive scenarios (e.g., very large databases) are not suitable to process big data, hence new approaches must be proposed. Here, the most relevant issues are: (*i*) exploiting data compression paradigms to improve querying performance, still ensuring the accuracy of answers; (*ii*) studying preference-based query processing solutions – how it copes with big data? (*iii*) scalability issues deriving from querying big data, and so forth.
- *Context-Aware Big Data Processing*. As highlighted before, the final goal of the SeDaSOMA framework consists in making available powerful data-intensive analytics to next-generation big data applications, which, in particular, exploit these analytics based on the fortunate Cloud computing paradigm. Following this conceptual vision, due to the massive amounts and the strongly heterogeneous nature of big data, it is mandatory to adopt context-aware approaches in order to "provide the right (big) data to the right application". Recently, this issue has become very relevant, even due to its strong connection with topics like big data advertisement and big data recommendation. In the context of the SeDaSOMA framework, this requirement is fulfilled by the Context-Aware Big Data Processing component. Here, the most relevant issues are: (*i*) context-aware algorithms and methods for big data; (*ii*) preference-based big data processing; (*iii*) scalable context-aware big data processing, and so forth.
- *Big Data Preparation for Analytics*. When designing and executing analytics over big data, a major challenge is big data preparation, i.e. how to pre-process big data in order to make analytics more effective and, possibly, more efficient. Indeed, this issue plays a critical role, due to the fact that data is characterized by extremely-massive size and strong heterogeneity. The lack of a "unifying" schema often makes analytics hard to be successful (e.g., in terms of "responsiveness"). The Big Data Preparation for Analytics component must answer questions such as how to enrich existing data? how to prune unnecessary data? how to normalize data? how to successfully extract from big data attributes and features that drive the analytics process? Moreover, since answers to some of those questions are only known during the analytics phase, the Big Data Preparation for Analytics component must provide on-demand preparation primitives in order to be able to respond to requests from the analytics one. Last but not least, all preparation primitives must be privacy-preserving especially when dealing with user data. We plan to refer to the SOLID ecosystem in order to manage and make sustainable the challenges of privacy [32].
- *Scalable Big Data Analytics*. Supporting scalable big data analytics is a major goal of the whole SeDaSOMA proposal. The responsible component, Scalable Big Data Analytics, deals with the issue of defining novel paradigms for next-generation analytics characterized by high responsiveness and high scalability. This component plays a critical role even because Cloud-aware big data vertical applications relying on top of the SeDaSOMA framework (and devoted to assess and possibly show the effectiveness and the reliability of the framework) are mainly based on it, and successfully exploit the results and the interactions of analytics to achieve their respective applicative goals. Here, relevant research issues are: (*i*) data-intensive analytics, i.e. analytics that interact with large-scale repositories like big data; (*ii*) declarative versus procedural analytics; (*iii*) quality-aware analytics – how to measure the "quality" of analytics? (*iv*) responsiveness issues of analytics; (*v*) scalability issues of analytics, and so forth.

The Big Data Analytics Layer plays the role of principal underlying layer for a collection of Cloud-aware big data vertical applications, whose main goal is to assess the effectiveness and the reliability of the SeDaSOMA framework, to which such applications are interfaced to. These applications adhere to the Cloud computing paradigm, and, particularly, they exploit the underlying analytics exposed by the SeDaSOMA framework in order to achieve their respective applicative goals, even in dependence on user/environment requirements.

In the SeDaSOMA framework, applications that are identified as critical fall in major areas of actual information society challenges drawn by the European data management guidelines. They are the following ones:

- *Social Big Data Management for Workplace Safety and Health*. The vertical application *Social Big Data Management for Workplace Safety and Health* focuses the attention on the problem of "serendipitily" discovering and integrating (big) data coming from multiple media sources such as blogs, social networks, emails, forums, communities, and so forth, in order to integrate and mine such data with the goal of sharing data among users as to fruitfully exploit them for workplace safety and health. In the so-delineated application scenario, people will have an "active" role, meaning that they can share their opinion, criticisms and positive/negative regards as to create a "real" (big-data-based) community with powerful analytics capabilities.

- Integration of (Big) Bank Data and (Big) Customer Data for Supporting Big Data Intelligence. The vertical application Integration of (Big) Bank Data and (Big) Customer Data for Supporting Big Data Intelligence deals with the problem of integrating bank data and customer data in order to improve the "on-line" experience of customers with typical banking services that may be delivered via different classes of devices. Both types of data (i.e., bank data and customer data) are big in nature, and they all expose the well-known 3V features of big data. The application will also integrate heterogeneous type of data, such as location data related to customers, and services, such as health services that may take advantage from their integration with bank data associated to customers, which are thus viewed as citizens by health services.
- *Big Data Advertisement on the Web for Opportunity Funding*. The vertical application *Big Data Advertisement on the Web for Opportunity Funding* considers the problems of using big data analytics (also combined with artificial intelligence algorithms) in order to support advanced Web marketing in a plethora of sectors (work, investments, etc.). This is a very exciting challenge for next-generation society, as being stirred-up by already well-consolidated technologies and platforms for the Web (e.g., *Google*, *Amazon*, *Alibaba*, etc.) and for social networks (e.g., *Facebook*, *Instagram*, *LinkedIn*, etc.) that allow providers and vendors to collect and store massive amounts of big data repositories that represent user profiles, user preferences, user goals, etc. On top of these big data repositories, artificial intelligence algorithms run in order to find the best (sub-optimal?) solution in a very wide family of Web application scenarios, such as buying, investing, dating, and so forth.

4 DISCUSSION

As we demonstrated through the paper, the issue of big data management and analytics is playing a critical role in the actual research community. The SeDaSOMA-framework's proposal is a direct consequence of this clear trend. SeDaSOMA includes many research innovations, among which the serendipitous paradigm is the major one.

What is next? In our vision, critical big data applications will more and more teach us what the novel requirements and features to capture will be, just like happened in other relevant cases (e.g., COVID-19 pandemic management and analytics [33]). On the other hand, *foundational and methodological aspects of big data management and analytics* still need to establish themselves largely, due to the fact that, on the application side, the evolution of big data management and analytics has been incredible fast during the last years.

5 CONCLUSIONS AND FUTURE WORK

By focusing on the emerging big data trend, in this paper we have introduced the anatomy of SeDaSOMA, a layered framework for supporting Se*rendipitous,* Da*ta-as-a-*S*ervice-oriented,* O*pen big data* M*anagement and* A*nalytics*. The proposed framework supports *advanced big data management and analytics* on several Cloud-aware application scenarios, including the "hot" ones: smart cities, social networks, sensor networks, intelligent transportation systems, and so forth. As to complete our contributions, we have also provided some example Cloud-aware big data vertical applications of SeDaSOMA in specific scenarios that are of great interest at now.

Future work is mainly oriented towards integrating our proposed framework with sustainable big data features, such as: (*i*) *visualization metaphors* (e.g., [34-35]), (*ii*) *flexible methodologies* (e.g., [36]), (*iii*) *uncertain management paradigms* (e.g., [37]), (*iv*) *intelligent data exchange approaches* (e.g., [38]).

ACKNOWLEDGMENTS

This research has been partially supported by the French PIA project "Lorraine Université d'Excellence", reference ANR-15-IDEX-04-LUE.

REFERENCES

[1] J. Chen, Y. Chen, X. Du, C. Li, J. Lu, S. Zhao, X. Zhou, "Big data challenge: a data management perspective". Frontiers Comput. Sci. 7(2), pp. 157-164, 2013
[2] H. Zhang, G. Chen, B.C. Ooi, K.-L. Tan, M. Zhang, "In-Memory Big Data Management and Processing: A Survey". IEEE Trans. Knowl. Data Eng. 27(7), pp. 1920-1948, 2015
[3] A. Siddiqa, I.A.T. Hashem, I. Yaqoob, M. Marjani, S. Shamshirband, A. Gani, F. Nasaruddin, "A survey of big data management: Taxonomy and state-of-the-art". J. Netw. Comput. Appl. 71, pp. 151-166, 2016
[4] P. Russom, "Big data analytics". TDWI Best Practices Report, fourth quarter, 2011
[5] P. Zikopoulos, C. Eaton, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data", McGraw-Hill Osborne Media, 1st. ed., 2011
[6] A. Cuzzocrea, C. De Maio, G. Fenza, V. Loia, M. Parente, "OLAP Analysis of Multidimensional Tweet Streams for Supporting Advanced Analytics". In: SAC 2016, pp. 992–999, 2016
[7] I.A.T. Hashem, V. Chang, N.B. Anuar, K.S. Adewole, I. Yaqoob, A. Gani, E. Ahmed, H. Chiroma, "The role of big data in smart city". Int. J. Inf. Manag. 36(5), pp. 748-758, 2016
[8] W. Tan, M.B. Blake, I. Saleh, S. Dustdar, "Social-Network-Sourced Big Data Analytics". IEEE Internet Comput. 17(5), pp. 62-69, 2013.
[9] S. Boubiche, D.E. Boubiche, A. Bilami, H. Toral-Cruz, "Big Data Challenges and Data Aggregation Strategies in Wireless Sensor Networks". IEEE Access 6, pp. 20558-20571, 2018
[10] L. Zhu, F.R. Yu, Y. Wang, B. Ning, T. Tang, "Big Data Analytics in Intelligent Transportation Systems: A Survey". IEEE Trans. on Intelligent Transportation Systems 20(1), pp. 383-398, 2019
[11] D. Laney, "3D Data Management: Controlling Data Volume, Velocity, and Variety". META Group Technical Report, 2001.
[12] A. Cuzzocrea, P. Serafino, "LCS-Hist: taming massive high-dimensional data cube compression". In: EDBT 2009, pp. 768-779, 2009.
[13] A. Cuzzocrea, D. Saccà, P. Serafino, "A Hierarchy-Driven Compression Technique for Advanced OLAP Visualization of Multidimensional Data Cubes". In: DaWaK 2006, pp. 106-119, 2006
[14] A. Cuzzocrea, "Improving range-sum query evaluation on data cubes via polynomial approximation". Data Knowl. Eng. 56(2), pp. 85-121, 2006
[15] A. Cuzzocrea, U. Matrangolo, "Analytical Synopses for Approximate Query Answering in OLAP Environments". In: DEXA 2004, pp. 359-370, 2004
[16] A Cuzzocrea, "Approximate OLAP query processing over uncertain and imprecise multidimensional data streams". In: DEXA 2013, pp. 156-173, 2013
[17] A. Cuzzocrea, R. Moussa, G. Xu, "OLAP*: Effectively and Efficiently Supporting Parallel OLAP over Big Data". In: MEDI 2013, pp. 38-49. 2013
[18] J. Han, "OLAP Mining: Integration of OLAP with Data Mining". In: DS-7 1997, pp. 3-20, 1997
[19] A. Cuzzocrea, "OLAP Intelligence: Meaningfully Coupling OLAP and Data Mining Tools and Algorithms". International Journal of Business Intelligence and Data Mining 4(3/4), pp. 213-218, 2009
[20] A. Cuzzocrea, "Scalable OLAP-based big data analytics over cloud infrastructures: Models, issues, algorithms". In: ICCBDC 2017, pp. 17-21, 2017
[21] A. Cuzzocrea, F. Jiang, C.K. Leung, D. Liu, A. Peddle, S.K. Tanbeer, "Mining popular patterns: a novel mining problem and its application to static transactional databases and dynamic data streams". Transactions on Large-Scale Data-and Knowledge-Centered Systems XXI, pp. 115-139, 2015

[22] J.J. Cameron, A. Cuzzocrea, F. Jiang, C.K. Leung, "Frequent pattern mining from dense graph streams". In: EDBT/ICDT Workshops 2014, pp. 240-247, 2014
[23] A. Adadi, "A survey on data-efficient algorithms in big data era". J. Big Data 8(1), pp. 1-54, 2021
[24] G. Chatzimilioudis, A. Cuzzocrea, D. Gunopulos, N. Mamoulis, "A novel distributed framework for optimizing query routing trees in wireless sensor networks via optimal operator placement". J. Comput. Syst. Sci. 79(3), pp. 349-368, 2013
[25] European Commission, Horizon 2020 – The EU Framework Programme for Research and Innovation, https://ec.europa.eu/programmes/horizon2020/en
[26] Y. Wang, J. Wei, M. Srivatsa, Y. Duan, W. Du, "IntegrityMR: Integrity assurance framework for big data analytics and management applications". In: BigData 2013, pp. 33-40, 2013
[27] S. Fiore, C. Palazzo, A. D'Anca, I.T. Foster, D.N. Williams, G. Aloisio, "A big data analytics framework for scientific data management". In: BigData 2013, pp. 1-8, 2013
[28] D. Puthal, S. Nepal, R. Ranjan, J. Chen, "A Secure Big Data Stream Analytics Framework for Disaster Management on the Cloud". In: HPCC/SmartCity/DSS 2016, pp. 1218-1225, 2016
[29] M.F. Abdullah, M. Ibrahim, H. Zulkifli, "Big Data Analytics Framework for Natural Disaster Management in Malaysia". In: IoTBDS 2017, pp. 406-411, 2017
[30] G. Terrazas, N. Ferry, S.M. Ratchev, "A cloud-based framework for shop floor big data management and elastic computing analytics". Comput. Ind. 109, pp. 204-214, 2019
[31] A. Jindal, N. Kumar, M. Singh, "A unified framework for big data acquisition, storage, and analytics for demand response management in smart cities". Future Gener. Comput. Syst. 108, pp. 921-934, 2020
[32] R. Buyle, R. Taelman, K. Mostaert, G. Joris, E. Mannens, R. Verborgh, T. Berners-Lee, "Streamlining governmental processes by putting citizens in control of their personal data". In: Proc. Int. Conf. on Electronic Governance and Open Society: Challenges in Eurasia 2019, pp. 346-359, 2019
[33] M.A. Elmeiligy, A.I. El-Desouky, S.M. El-Ghamrawy, "A Multi-Dimensional Big Data Storing System for Generated COVID-19 Large-Scale Data using Apache Spark". CoRR abs/2005.05036, 2020
[34] A. Cuzzocrea, S. Mansmann, "OLAP Visualization". Encyclopedia of Data Warehousing and Mining 2009, pp. 1439-1446, 2009
[35] K.E. Barkwell, A. Cuzzocrea, C.K. Leung, A.A. Ocran, J.M. Sanderson, "Big data visualisation and visual analytics for music data mining", In: IV 2018, pp. 235-240, 2018
[36] D. Teng, J. Kong, F. Wang, "Scalable and flexible management of medical image big data". Distributed Parallel Databases 37(2), pp. 235-250, 2019
[37] A.K. Shukla, P.K. Muhuri, "Big-data clustering with interval type-2 fuzzy uncertainty modeling in gene expression datasets". Eng. Appl. Artif. Intell. 77, pp. 268-282, 2019
[38] Y. Chen, J. Guo, C. Li, W. Ren, "FaDe: A Blockchain-Based Fair Data Exchange Scheme for Big Data Sharing". Future Internet 11(11), art. 225, 2019