Big Data and Cloud Computing: A Survey of the State-of-the-Art and Research Challenges

Georgios Skourletopoulos, Constandinos X. Mavromoustakis, George Mastorakis, Jordi Mongay Batalla, Ciprian Dobre, Spyros Panagiotakis and Evangelos Pallis

Abstract The proliferation of data warehouses and the rise of multimedia, social media and the Internet of Things (IoT) generate an increasing volume of structured, semi-structured and unstructured data. Towards the investigation of these large volumes of data, big data and data analytics have become emerging research fields, attracting the attention of the academia, industry and governments. Researchers, entrepreneurs, decision makers and problem solvers view 'big data' as the tool to revolutionize various industries and sectors, such as business, healthcare, retail, research, education and public administration. In this context, this survey chapter

G. Skourletopoulos (✉) · C.X. Mavromoustakis
Mobile Systems (MoSys) Laboratory, Department of Computer Science, University of Nicosia, Nicosia, Cyprus
e-mail: skourletopoulos.g@unic.ac.cy

C.X. Mavromoustakis
e-mail: mavromoustakis.c@unic.ac.cy

G. Mastorakis · S. Panagiotakis · E. Pallis
Department of Informatics Engineering, Technological Educational Institute of Crete, Heraklion, Crete, Greece
e-mail: gmastorakis@staff.teicrete.gr

S. Panagiotakis
e-mail: spanag@teicrete.gr

E. Pallis
e-mail: pallis@pasiphae.eu

J.M. Batalla
Warsaw University of Technology, Nowowiejska Str. 15/19, Warsaw, Poland
e-mail: jordim@interfree.it

C. Dobre
Faculty of Automatic Control and Computers, Department of Computer Science, University Politehnica of Bucharest, Bucharest, Romania
e-mail: ciprian.dobre@cs.pub.ro

C.X. Mavromoustakis et al. (eds.), *Advances in Mobile Cloud Computing and Big Data in the 5G Era*, Studies in Big Data 22,
DOI 10.1007/978-3-319-45145-9_2

presents a review of the current big data research, exploring applications, opportunities and challenges, as well as the state-of-the-art techniques and underlying models that exploit cloud computing technologies, such as the big data-as-a-service (BDaaS) or analytics-as-a-service (AaaS).

Keywords Big data · Data analytics · Data management · Big data-as-a-service · Analytics-as-a-service · Business intelligence · Lease storage · Cloud computing · Cost-benefit analysis model

1 Introduction

The Internet penetration constantly increases, as more and more people browse the Web, use email and social network applications to communicate with each other or access wireless multimedia services, such as mobile TV [27, 43]. Additionally, several demanding mobile network services are now available, which require increased data rates for specific operations, such as device storage synchronization to cloud computing servers or high resolution video [34–36]. The access to such a global information and communication infrastructure along with the advances in digital sensors and storage have created very large amounts of data, such as Internet, sensor, streaming or mobile device data. Additionally, data analysis is the basis for investigations in many fields of knowledge, such as science, engineering or management. Unlike web-based big data, location data is an essential component of mobile big data, which are harnessed to optimize and personalize mobile services. Hence, an era where data storage and computing become utilities that are ubiquitously available is now introduced.

Furthermore, algorithms have been developed to connect datasets and enable more sophisticated analysis. Since innovations in data architecture are on our doorstep, the 'big data' paradigm refers to very large and complex data sets (i.e., petabytes and exabytes of data) that traditional data processing systems are inadequate to capture, store and analyze, seeking to glean intelligence from data and translate it into competitive advantage. As a result, big data needs more computing power and storage provided by cloud computing platforms. In this context, cloud providers, such as IBM [23], Google [17], Amazon [2] and Microsoft [38], provide network-accessible storage priced by the gigabyte-month and computing cycles priced by the CPU-hour [8].

Although big data is still in the preliminary stages, comprehensive surveys exist in the literature [1, 9–11, 20, 37, 59]. This survey article aims at providing a holistic perspective on big data and big data-as-a-service (BDaaS) concepts to the research community active on big data-related themes, including a critical revision of the current state-of-the-art techniques, definition and open researches issues. Following this introductory section, Sect. 2 presents related work approaches in the literature, including the architecture and possible impact areas. Section 3 demonstrates the business value and long-term benefits of adopting big data-as-a-service business models and attempts to communicate the findings to non-technical stakeholders, while Sect. 4 points out opportunities, challenges and open research issues in the big data domain. Finally, Sect. 5 concludes this tutorial chapter.

2 Big Data: Background and Architecture

IBM data scientists argue that the key dimensions of big data are the “4Vs”: volume, velocity, variety and veracity [21]. As large and small enterprises constantly attempt to design new products to deal with big data, the open source platforms, such as Hadoop [53], give the opportunity to load, store and query a massive scale of data and execute advanced big data analytics in parallel across a distributed cluster. Batch-processing models, such as MapReduce [14], enable the data coordination, combination and processing from multiple sources. Many big data solutions in the market exploit external information from a range of sources (e.g., social networks) for modelling and sentiment analysis, such as the IBM Social Media Analytics Software as a Service solution [22]. Cloud providers have already begun to establish new data centers for hosting social networking, business, media content or scientific applications and services. In this direction, the selection of the data warehouse technology depends on several factors, such as the volume of data, the speed with which the data is needed or the kind of analysis to be performed [25]. A conceptual big data warehouse architecture is presented in Fig. 1 [24].

Another significant challenge is the delivery of big data capabilities through the cloud. The adoption of big data-as-a-service (BDaaS) business models enables the effective storage and management of very large data sets and data processing from an outside provider, as well as the exploitation of a full range of analytics capabilities (i.e., data and predictive analytics or business intelligence are provided as service-based applications in the cloud). In this context, Zheng et al. [59] critically review the service-generated big data and big data-as-a-service (see Fig. 2) towards

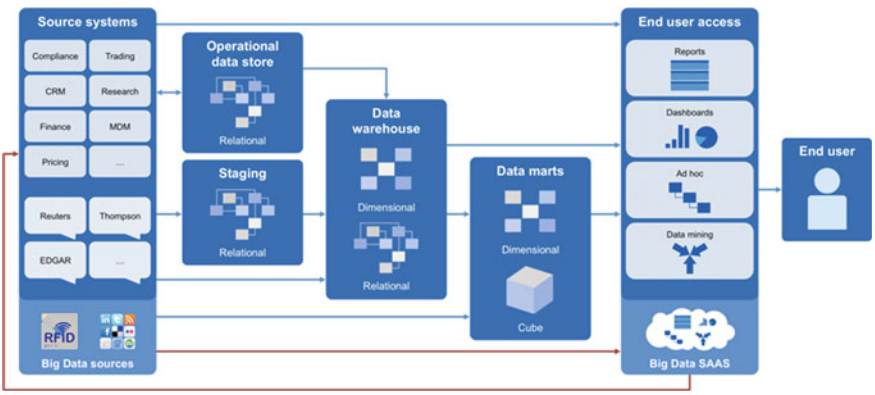

Fig. 1 A conceptual big data warehouse architecture

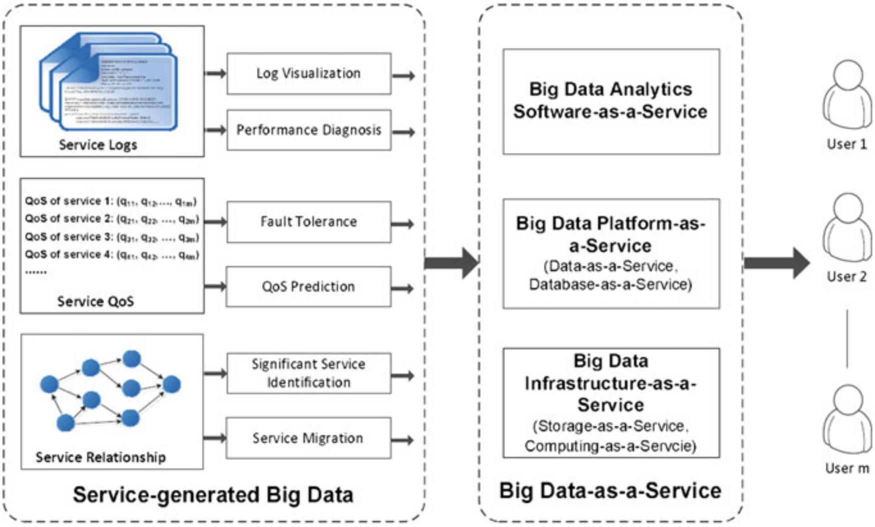

Fig. 2 Service-generated big data and big data-as-a-service as presented by Zheng et al. [59]

the proposal of an infrastructure to provide functionality for managing and analyzing different types of service-generated big data. A big data-as-a-service framework has been also employed to provide big data services and data analytics results to users, enhance efficiency and reduce cost.

The development of a cloud-supported big data mining platform, which provides statistical and data analytics functions, has been also explored [56]. In this research work, the platform's architecture is composed of four layers (i.e., infrastructure, virtualization, data set processing and services), implementing the K-means algorithm. A big data analytics-related platform was proposed by Park et al. [40], which includes a CCTV metadata analytics service and aims to manage big data and develop analytics algorithms through collaboration between data owners, scientists and developers. Since modern enterprises request new solutions for enterprise data warehousing (EDW) and business intelligence (BI), a big data provisioning solution was elaborated by Vaquero et al. [55], combining hierarchical and peer-to-peer data distribution techniques to reduce the data loading time into the virtual machines (VMs). The proposed solution includes dynamic topology and software configuration management techniques for better quality of experience (QoE) and achieves to reduce the setup time of virtual clusters for data processing in the cloud. A cloud-based big data analytics service provisioning platform, named CLAaaS, has been presented in the literature along with a taxonomy to identify significant features of the workflow systems, such as multi-tenancy for a wide range of analytic tools and back-end data sources, user group customization and web collaboration [60]. An overview of the analytics workflow for big data is shown in Fig. 4 [3]. On the other hand, an admission control and resource scheduling algorithm is examined in another work [58], which manages to satisfy the quality of service requirements

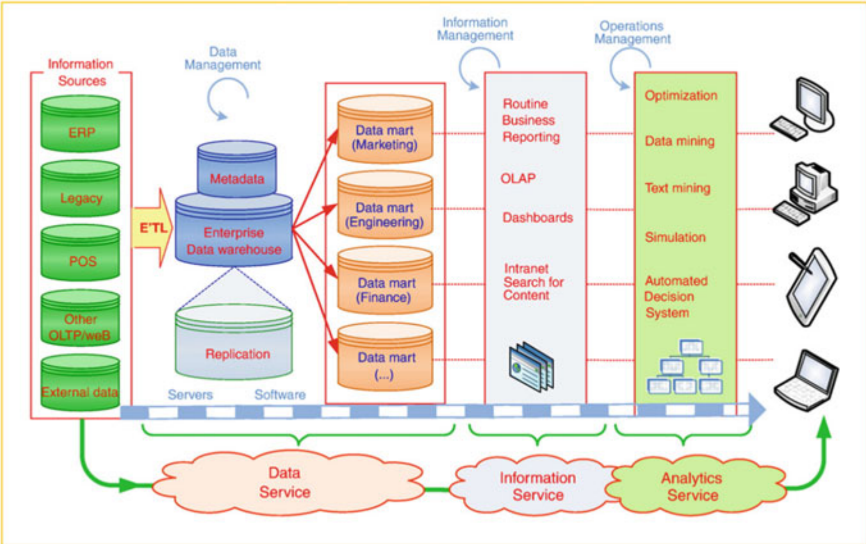

Fig. 3 A conceptual architecture of service-oriented decision support systems as presented by Demirkan and Delen [15]

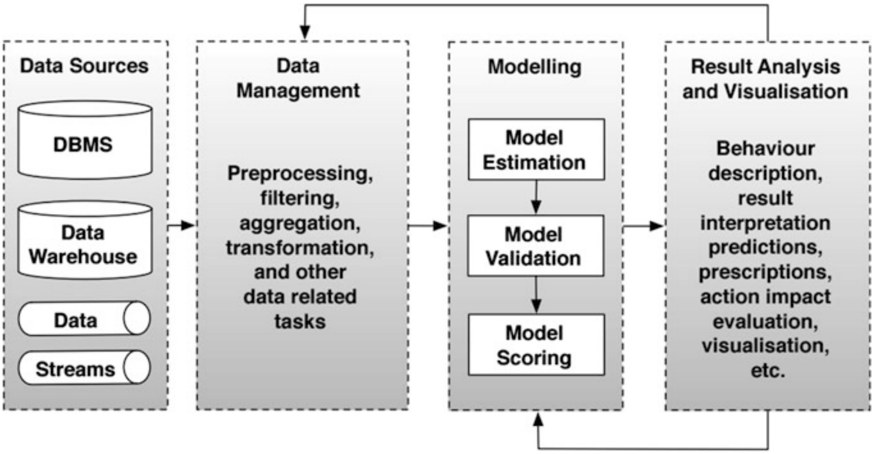

Fig. 4 Analytics workflow for big data as presented by Assunção et al. [3]

of requests, adhering to the Service Level Agreements (SLAs) guarantees, and improve the Analytics-as-a-Service (AaaS) providers' competitiveness and profitability. A framework for service-oriented decision support systems (DSS) in the cloud has been also investigated, focusing on the product-oriented decision support systems environment and exploring engineering-related issues [15]. A conceptual architecture of service-oriented decision support systems is shown in Fig. 3.

The growth of cloud computing, big data and analytics [52] compels businesses to turn into big data-as-a-service solutions in order to overcome common challenges, such as data storage or processing power. Although there is related work in the literature in the general area of cost-benefit analysis in cloud and mobile cloud computing environments, a research gap is observed towards the evaluation and classification of big data-as-a-service business models. Several research efforts have been devoted comparing the monetary cost-benefits of cloud computing with desktop grids [26], examining cost-benefit approaches of using cloud computing to extend the capacity of clusters [13] or calculating the cloud total cost of ownership and utilization cost [30] to evaluate the economic efficiency of the cloud. Finally, novel metrics for predicting and quantifying the technical debt on cloud-based software engineering and cloud-based service level were also proposed in the literature from the cost-benefit viewpoint [44, 45] and extended evaluation results are discussed by Skourletopoulos et al. [46].

3 Cloud-Supported Big Data: Towards a Cost-Benefit Analysis Model in Big Data-as-a-Service (BDaaS)

In previous research works, the cloud was considered as a marketplace [7], where the storage and computing capabilities of the cloud-based system architectures can be leased off [47, 49, 50]. Likewise, the rise of large data centers has created new business models, where businesses lease storage in a pay-as-you-go service-oriented manner [32, 57]. In this direction, the big data-as-a-service (BDaaS) model was introduced in order to provide common big data services, boost efficiency and reduce cost [51]. Communicating the business value and long-term benefits of adopting big data-as-a-service business models against the conventional high-performance data warehouse appliances to non-technical stakeholders is imperative. In this book chapter, a brief survey of a novel quantitative, cloud-inspired cost-benefit analysis metric in big data-as-a-service is presented, based on previous research studies in the literature [48]. Hence, the cost analysis (CA) modelling from the conventional data warehouse appliance (DWH) viewpoint takes the following form and the descriptions of the exploited variables are shown in Table 1:

$$CA_i = 12 * (C_{s/m} * S_{max}), \quad i \geq 1 \ and \ S_{curr} \leq S_{max} \tag{1}$$

where,

$$C_{s/m} = C_{s/m(max)} = C_{\alpha/m(max)} + C_{\gamma/m(max)} + C_{\eta/m(max)} + C_{\theta/m(max)} + C_{\kappa/m(max)} + C_{\lambda/m(max)} + C_{\mu/m(max)} + C_{\sigma/m(max)}$$

As the benefits of cloud computing (i.e., scalability) do not stand in data warehouse appliances, the cost analysis approach adopted in this study does not consider the storage capacity currently used (S_{curr}). Therefore, the cost variations,

Table 1 Notations and variable descriptions

Symbol	Variable description
CA	The cost analysis calculation results, which are expressed in monetary units
C_D	The benefits calculation results, which are expressed in monetary units
i	The index of the year
$C_{s/m}$	The initial monthly cost for leasing cloud storage, which is expressed in monetary units
S_{max}	The maximum storage capacity
S_{curr}	The storage currently used
Δ_0	The cost formation for leasing cloud storage regarding the second year of the period of l-years, once the corresponding variation in the monthly cost is applied, which is expressed in monetary units
δ_1	The total variation regarding the cost for leasing cloud storage for the second year of the period of l-years, which is represented by a percentage value
Δ_i	The cost formation for leasing cloud storage from the third year and onwards, once the corresponding variation in the monthly cost is applied, which is expressed in monetary units
δ_i	The total variation regarding the cost for leasing cloud storage from the third year and onwards, which is represented by a percentage value
B_0	The storage used during the second year, once the corresponding variation in the demand is applied
β_1	The variation in the demand for storage capacity regarding the second year, which is represented by a percentage value
B_i	The storage used from the third year and onwards, once the corresponding variation in the demand is applied
β_i	The variation in the demand for storage capacity from the third year and onwards, which is represented by a percentage value
C_α	The data storage cost
C_γ	The document storage cost
C_η	The maintenance services cost
C_θ	The network cost
C_κ	The on-demand I/O cost
C_λ	The operations cost
C_μ	The server cost
C_σ	The technical support cost
$\alpha_i\%$	The variation in the monthly data storage cost, which is represented by a percentage value
$\gamma_i\%$	The variation in the monthly document storage cost, which is represented by a percentage value
$\eta_i\%$	The variation in the monthly maintenance services cost, which is represented by a percentage value
$\theta_i\%$	The variation in the monthly network cost, which is represented by a percentage value
$\kappa_i\%$	The variation in the monthly on-demand I/O cost, which is represented by a percentage value

(continued)

Table 1 (continued)

Symbol	Variable description
$\lambda_i\%$	The variation in the monthly operations cost, which is represented by a percentage value
$\mu_i\%$	The variation in the monthly server cost, which is represented by a percentage value
$\sigma_i\%$	The variation in the monthly technical support cost, which is represented by a percentage value

due to the fluctuations in the demand for storage capacity, do not apply as long as $S_{curr} \leq S_{max}$ and the true benefits are always zero ($C_D = 0$) over the years. In case of such an increase in the demand for storage capacity that $S_{curr} > S_{max}$, incremental capacity should be added to the storage systems with overhead and downtime. On the contrary, the cost-benefit analysis modelling from the big data-as-a-service point of view takes the following form during the first year (i.e., Eqs. 2 and 4) and from the second year and onwards (i.e., Eqs. 3 and 5):

$$CA_1 = 12 * (C_{s/m} * S_{curr}) \tag{2}$$

$$CA_i = 12 * (\Delta_{i-2} * B_{i-2}), i \geq 2 \tag{3}$$

$$C_{D_1} = 12 * [C_{s/m} * (S_{max} - S_{curr})] \tag{4}$$

$$C_{D_i} = 12 * [\Delta_{\iota-2} * (S_{max} - B_{i-2})], i \geq 2 \tag{5}$$

where,

$$C_{s/m} = C_{s/m(curr)} = C_{\alpha/m(curr)} + C_{\gamma/m(curr)} + C_{\eta/m(curr)} + C_{\theta/m(curr)} + C_{\kappa/m(curr)} + C_{\lambda/m(curr)} + C_{\mu/m(curr)} + C_{\sigma/m(curr)}$$

$$\Delta_0 = (1 + \delta_1\,\%) * C_{s/m}$$

$$\Delta_i = (1 + \delta_{i+1}\,\%) * \Delta_{i-1}, i \geq 1$$

$$\delta_i\,\% = a_i\,\% + \gamma_i\,\% + \eta_i\,\% + \theta_i\,\% + \kappa_i\,\% + \lambda_i\,\% + \mu_i\,\% + \sigma_i\,\%, i \geq 1$$

$$B_0 = (1 + \beta_1\,\%) * S_{curr}$$

$$B_i = (1 + \beta_{i+1}\,\%) * B_{i-1}, i \geq 1$$

The amount of profit not earned due to the underutilization of the storage capacity is measured, under the assumption that fluctuations in the demand for cloud storage occur. The possible upgradation of the storage and the risk of entering into new and accumulated costs in the future are also examined. The cloud storage capacity to be leased off is evaluated with respect to the following assumptions:

- The cloud storage is subscription-based and the billing vary over the period of l-years due to the fluctuations in the demand for storage capacity (i.e., gigabyte per month).
- The total network cost consists of bandwidth usage, egress and data transfer costs between regional and multi-region locations. As the cloud-based, always-on mobile services are usually sensitive to network bandwidth and latency [42], the additional network cost is expected to satisfy the outbound network traffic demands in order to avoid delays.
- Since the content retrieval from a bucket should be faster than the default, the additional on-demand I/O cost enables to increase the throughput [4, 39].
- The additional server cost stems from the additional CPU cores and the amount of memory required for processing.

Two possible types of benefits calculation results are encountered, when leasing cloud storage:

- Positive calculation results, which point out the underutilization of the storage capacity.
- Negative calculation results, which reveal the immediate need for upgradation. This need stimulates additional costs; however, the total amount of accumulated cost in conventional data warehouse appliances is not comparable, as the earnings by adopting a big data-as-a-service business model can be reinvested on the additional storage required, maximizing the return on investment.

Towards the evaluation of big data-as-a-service business models and the increase in the return on investment, the way the benefits overcome the costs is of significant importance [12, 28, 32, 41]. An illustrative example emphasizes on the need to consolidate data from different sources. Cost analysis and benefits comparisons are performed during a 5-year period of time ($l = 5$) prior to adoption of either a conventional data warehouse or a big data-as-a-service business model. The predicted variations in the demand for cloud storage with respect to two case scenarios are shown in Table 2.

In this framework, the first case scenario reveals that adopting a big data-as-a-service business model is more cost-effective than a conventional data warehouse, as the cost analysis results for the big data-as-a-service model have the least positive values throughout the 5-year period. The benefits calculation results are positive in big data-as-a-service business models, while the benefits results are always zero in conventional data warehouse business models (Figs. 5 and 6).

Table 2 Variations in the demand for cloud storage regarding two case scenarios

Term	Case scenario 1	Case scenario 2
Year 1–2	$\beta_{1\,\%} = 5\ \%$	$\beta_{1\,\%} = 10\ \%$
Year 2–3	$\beta_{2\,\%} = 15\ \%$	$\beta_{2\,\%} = 22\ \%$
Year 3–4	$\beta_{3\,\%} = 20\ \%$	$\beta_{3\,\%} = 35\ \%$
Year 4–5	$\beta_{4\,\%} = 23\ \%$	$\beta_{4\,\%} = 40\ \%$

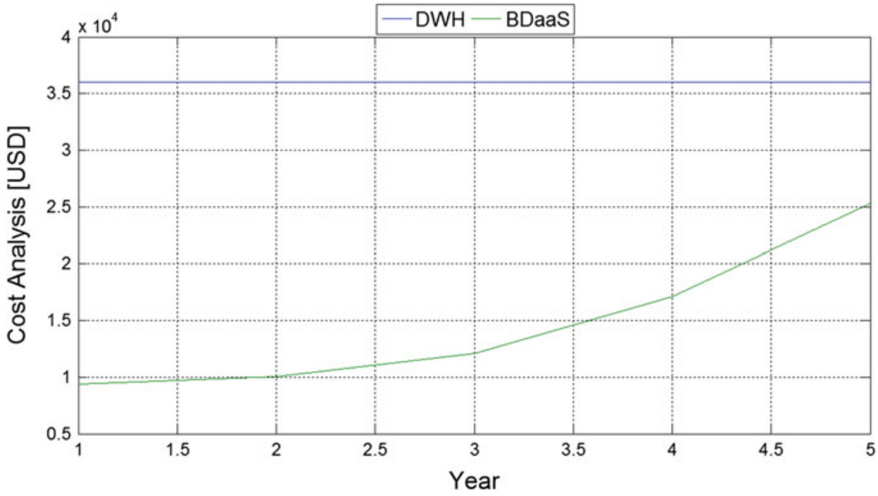

Fig. 5 Cost analysis for the first case scenario

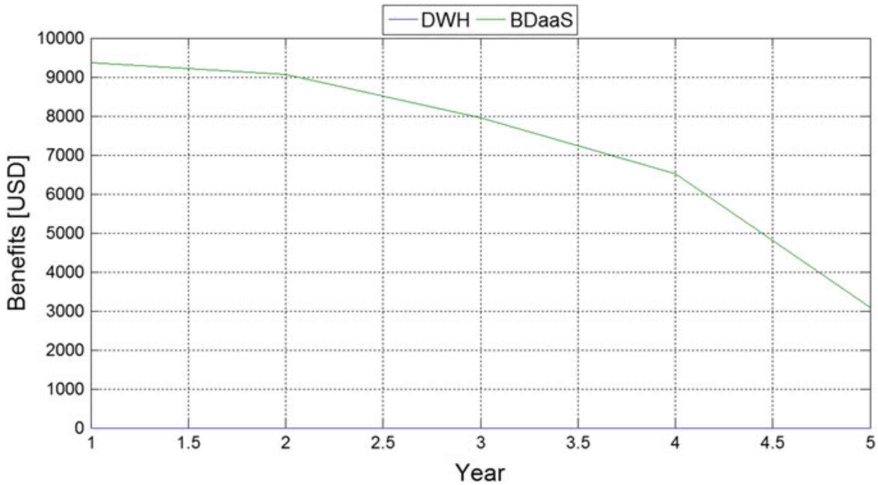

Fig. 6 Benefits analysis for the first case scenario

On the other hand, the second case scenario points out the cost-effectiveness and the benefits gained by adopting the big data-as-a-service model during the first four years. However, the benefits calculations results become negative during the fifth year, indicating the need for immediate upgradation to meet the demand requirements. The necessity for upgradation is also witnessed at the increased costs compared to those in the traditional data warehouse approach. In this direction, the earnings gained throughout the period, due to the selection of the dig data-as-a-service business model, will be reinvested on the additional storage required, maximizing the return on investment (ROI) (Figs. 7 and 8).

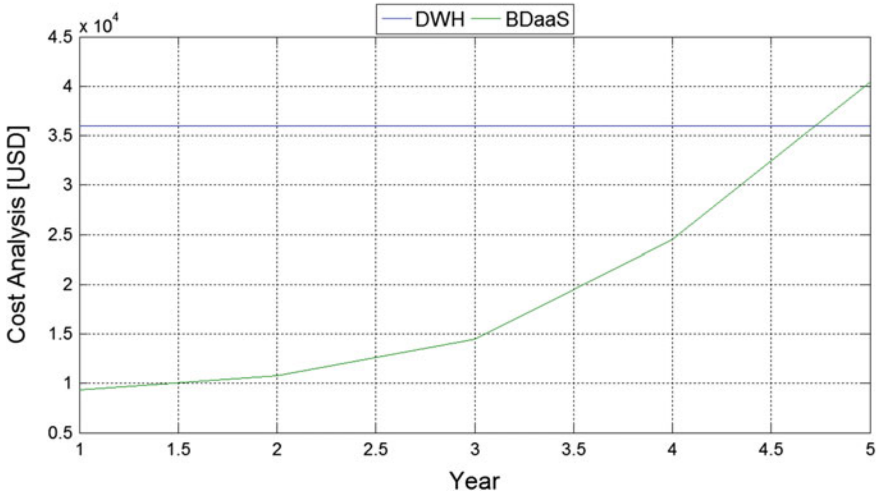

Fig. 7 Cost analysis for the second case scenario

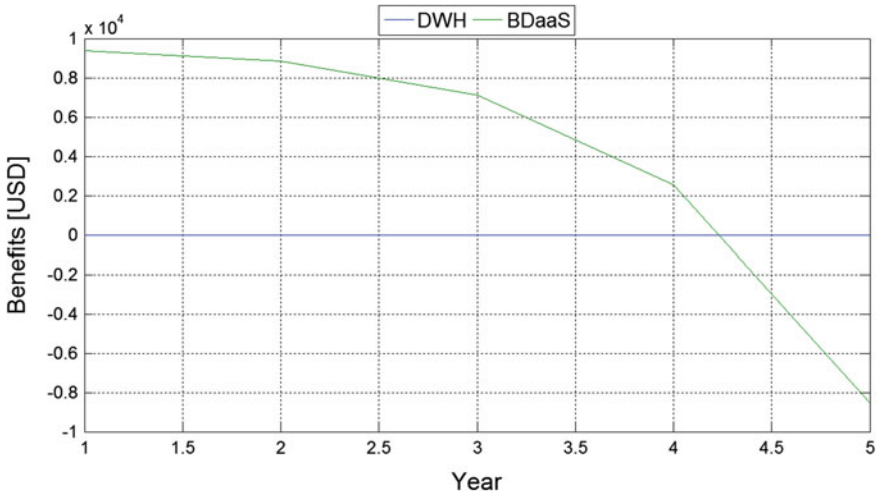

Fig. 8 Benefits analysis for the second case scenario

4 Challenges and Open Research Issues

The rise and development of social networks, multimedia, electronic commerce (e-Commerce) and cloud computing have increased considerably the data. Additionally, since the needs of enterprise analytics are constantly growing, the conventional hub-and-spoke architectures cannot satisfy the demands and, therefore, new and enhanced architectures are necessary [15]. In this context, new challenges and open research issues are encountered, including storage, capture, processing, filtering, analysis, curation, search, sharing, visualization, querying and privacy of the very large volumes of data. The aforementioned issues are categorized and elaborated as follows [11]:

- **Data storage and management**: Since big data are dependent on extensive storage capacity and data volumes grow exponentially, the current data management systems cannot satisfy the needs of big data due to limited storage capacity. In addition, the existing algorithms are not able to store data effectively because of the heterogeneity of big data.
- **Data transmission and curation**: Since network bandwidth capacity is the major drawback in the cloud, data transmission is a challenge to overcome, especially when the volume of data is very large. For managing large-scale and structured datasets, data warehouses and data marts are good approaches. Data warehouses are relational database systems that enable the data storage, analysis and reporting, while the data marts are based on data warehouses and facilitate the analysis of them. In this context, NoSQL databases [19] were introduced as a potential technology for large and distributed data management and database design. The major advantage of NoSQL databases is the schema-free orientation, which enables the quick modification of the structure of data and avoids rewriting the tables.
- **Data processing and analysis**: Query response time is a significant issue in big data, as adequate time is needed when traversing data in a database and performing real-time analytics. A flexible and reconfigured grid along with the big data preprocessing enhancement and consolidation of application- and data-parallelization schemes can be more effective approaches for extracting more meaningful knowledge from the given data sets.
- **Data privacy and security**: Since the host of data or other critical operations can be performed by third-party services or infrastructures, security issues are witnessed with respect to big data storage and processing. The current technologies used in data security are mainly static data-oriented, although big data entails dynamic change of current and additional data or variations in attributes. Privacy-preserving data mining without exposing sensitive personal information is another challenging field to be investigated.

5 Summary and Conclusion

Since networking is ubiquitous and vast amounts of data are now available, big data is envisioned to be the tool for productivity growth, innovation and consumer surplus. Huge opportunities related to advanced big data analytics and business intelligence are at the forefront of research, focusing on the investigation of innovative business-centric methodologies that can transform various sectors and industries, such as e-commerce, market intelligence, e-government, healthcare and security [4, 29, 31, 54]. To this end, this tutorial paper discusses the current big data research and points out the research challenges and opportunities in this field by exploiting cloud computing technologies and building new models [5, 6, 16, 18, 33–36]. A cost-benefit analysis is also performed towards measuring the long-term benefits of adopting big data-as-a-service business models in order to support data-driven decision making and communicate the findings to non-technical stakeholders.

Acknowledgements The authors would like to thank the anonymous reviewers for their constructive comments and insights on the manuscript. Their suggestions have contributed greatly to the high quality and improvement of this article. The authors would also like to acknowledge networking support by the ICT COST Action IC1303: Algorithms, Architectures and Platforms for Enhanced Living Environments (AAPELE) and the ICT COST Action IC1406 “High-Performance Modelling and Simulation for Big Data Applications” (cHiPSet).

References

1. Agrawal, D., Das, S., El Abbadi, A.: Big data and cloud computing: current state and future opportunities. In: Proceedings of the 14th International Conference on Extending Database Technology (EDBT/ICDT’11), pp. 530–533 (2011)
2. Amazon Web Services, Inc.: Elastic Compute Cloud (EC2). http://aws.amazon.com/ec2 (2015). Accessed 18 Oct 2015
3. Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A.S., Buyya, R.: Big data computing and clouds: trends and future directions. J. Parallel Distrib. Comput. **79–80**, 3–15 (2015)
4. Batalla, J.M., Kantor, M., Mavromoustakis, C.X., Skourletopoulos, G., Mastorakis, G.: A novel methodology for efficient throughput evaluation in virtualized routers. In: Proceedings of the IEEE International Conference on Communications (ICC 2015)—Communications Software, Services and Multimedia Applications Symposium (CSSMA), London, UK, pp. 6899–6905 (2015)
5. Batalla, J.M., Mavromoustakis, C.X., Mastorakis, G., Sienkiewicz, K.: On the track of 5G radio access network for IoT wireless spectrum sharing in device positioning applications. In: Internet of Things (IoT) in 5G Mobile Technologies, pp. 25–35. Springer International Publishing (2016)
6. Batalla, J.M., Mastorakis, G., Mavromoustakis, C.X., Zurek, J.: On cohabitating networking technologies with common wireless access for Home Automation Systems purposes. Special Issue on “Enabling Wireless Communication and Networking Technologies for the Internet of Things”. IEEE Wirel. Commun. Mag. (2016)
7. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. Future Gener. Comput. Syst. **25**, 599–616 (2009)
8. Buyya, R., Ranjan, R., Calheiros, R.N.: Intercloud: utility-oriented federation of cloud computing environments for scaling of application services. Algorithms Arch. Parallel Process. **6081**, 13–31 (2010)
9. Chen, C.L.P., Zhang, C.Y.: Data-intensive applications, challenges, techniques and technologies: a survey on big data. Inf. Sci. **275**, 314–347 (2014)

10. Chen, H., Chiang, R.H.L., Storey, V.C.: Business intelligence and analytics: from big data to big impact. MIS Q. **36**, 1165–1188 (2012)
11. Chen, M., Mao, S., Liu, Y.: Big data: a survey. Mobile Netw. Appl. **19**, 171–209 (2014)
12. Ciobanu, R.-I., Marin, R.-C., Dobre, C., Cristea, V., Mavromoustakis, C.X., Mastorakis, G.: Opportunistic dissemination using context-based data aggregation over interest spaces. In: Proceedings of IEEE International Conference on Communications 2015 (IEEE ICC 2015), London, UK, 08–12 June 2015
13. De Assunção, M.D., Di Costanzo, A., Buyya, R.: A cost-benefit analysis of using cloud computing to extend the capacity of clusters. Clust. Comput. **13**, 335–347 (2010)
14. Dean, J., Ghemawat, S.: Mapreduce: simplified data processing on large clusters. Commun. ACM **51**, 107–113 (2008)
15. Demirkan, H., Delen, D.: Leveraging the capabilities of service-oriented decision support systems: putting analytics and big data in cloud. Decis. Support Syst. **55**, 412–421 (2013)
16. Goleva, R., Stainov, R., Wagenknecht-Dimitrova, D., Mirtchev, S., Atamian, D., Mavromoustakis, C.X., Mastorakis, G., Dobre, C., Savov, A., Draganov, P.: Data and traffic models in 5G network. In: Internet of Things (IoT) in 5G Mobile Technologies, pp. 485–499. Springer International Publishing (2016)
17. Google, Inc.: App engine—platform as a service. https://cloud.google.com/appengine (2015). Accessed 18 Oct 2015
18. Hadjioannou, V., Mavromoustakis, C.X., Mastorakis, G., Batalla, J.M., Kopanakis, I., Perakakis, E., Panagiotakis, S.: Security in smart grids and smart spaces for smooth IoT deployment in 5G. In: Internet of Things (IoT) in 5G Mobile Technologies, pp. 371–397. Springer International Publishing (2016)
19. Han, J., Haihong, E., Le, G., Du, J.: Survey on NoSQL database. In: Proceedings of the 2011 6th International Conference on Pervasive Computing and Applications (ICPCA), Port Elizabeth, pp. 363–366 (2011)
20. Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U.: The rise of "big data" on cloud computing: review and open research issues. Inf. Syst. **47**, 98–115 (2015)
21. IBM Corporation: IBM big data & analytics hub: the four V's of big data. http://www.ibmbigdatahub.com/infographic/four-vs-big-data (2014). Accessed 18 Oct 2015
22. IBM Corporation: IBM social media analytics software as a service. http://www-03.ibm.com/software/products/en/social-media-analytics-saas (2015a). Accessed 18 October 2015
23. IBM Corporation: IBM bluemix. http://www.ibm.com/cloud-computing/bluemix (2015b). Accessed 18 Oct 2015
24. Informatica Corporation: Making sense of big data. https://marketplace.informatica.com/solutions/making_sense_of_big_data (2015). Accessed 18 Oct 2015
25. Inmon, W.H.: Building the Data Warehouse. Wiley (2005)
26. Kondo, D., Javadi, B., Malecot, P., Cappello, F., Anderson, D.P.: Cost-benefit analysis of cloud computing versus desktop grids. In: Proceedings of the 2009 IEEE International Symposium on Parallel and Distributed Processing (IPDPS 2009), Rome, pp. 1–12 (2009)
27. Kryftis, Y., Mavromoustakis, C.X., Batalla, J.M., Mastorakis, G., Pallis, E., Skourletopoulos, G.: Resource usage prediction for optimal and balanced provision of multimedia services. Proceedings of the 2014 IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD 2014), pp. 255–259. Greece, Athens (2014)
28. Kryftis, Y., Mavromoustakis, C.X., Mastorakis, G., Pallis, E., Batalla, J.M., Rodrigues, J., Dobre, C., Kormentzas, G.: Resource usage prediction algorithms for optimal selection of multimedia content delivery methods. In: Proceedings of IEEE International Conference on Communications 2015 (IEEE ICC 2015), London, UK, 08–12 June 2015
29. Kryftis, Y., Mastorakis, G., Mavromoustakis, C.X., Batalla, J.M., Rodrigues, J., Drobre, C.: Resource usage prediction models for optimal multimedia content provision. IEEE Syst. J. (2016)

30. Li, X., Li, Y., Liu, T., Qiu, J., Wang, F.: The method and tool of cost analysis for cloud computing. In: 2009 IEEE International Conference on Cloud Computing (CLOUD'09), Bangalore, pp. 93–100 (2009)
31. Markakis, E., Mastorakis, G., Negru, D., Pallis, E., Mavromoustakis, C.X., Bourdena, A.: A context-aware system for efficient peer-to-peer content provision. In: Dobre, C., Xhafa, F. (eds.) Pervasive Computing: Next Generation Platforms for Intelligent Data Collection. Elsevier (2016)
32. Mastorakis, G., Markakis, E., Pallis, E., Mavromoustakis, C.X., Skourletopoulos, G.: Virtual network functions exploitation through a prototype resource management framework. Proceedings of the 2014 IEEE 6th International Conference on Telecommunications and Multimedia (TEMU 2014), pp. 24–28. Heraklion, Crete, Greece (2014)
33. Mavromoustakis, C.X., Dimitriou, C., Mastorakis, G., Pallis, E.: Real-time performance evaluation of F-BTD scheme for optimized QoS energy conservation in wireless devices. In: Proceedings of IEEE Globecom 2013, 2nd IEEE Workshop on Quality of Experience for Multimedia Communications (QoEMC2013), Atlanta, GA, USA, 09–13 Dec 2013
34. Mavromoustakis, C.X., Mastorakis, G., Bourdena, A., Pallis, E., Kormentzas, G., Rodrigues, J.J.P.C.: Context-oriented opportunistic cloud offload processing for energy conservation in wireless devices. Proceedings of the IEEE Global Communications Conference (GLOBECOM 2014)—The Second International Workshop on Cloud Computing Systems, Networks, and Applications (CCSNA), pp. 24–30. Austin, Texas, USA (2014a)
35. Mavromoustakis, C.X., Andreou, A., Mastorakis, G., Bourdena, A., Batalla, J.M., Dobre, C.: On the performance evaluation of a novel offloading-based energy conservation mechanism for wireless devices. In: Proceedings of the 6th International Conference on Mobile Networks and Management (MONAMI 2014), 22–24 Sept 2014. Wuerzburg, Germany (2014b)
36. Mavromoustakis, C.X., Mastorakis, G., Bourdena, A., Pallis, E., Kormentzas, G., Dimitriou, C.: Joint energy and delay-aware scheme for 5G mobile cognitive radio networks. In: Proceedings of IEEE GlobeCom 2014, Symposium on Selected Areas in Communications: GC14 SAC Green Communication Systems and Networks, Austin, TX, USA (2014c)
37. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J., Barton, D., Court, D.: Big data: the management revolution. Harv. Bus. Rev. 59–68 (2012)
38. Microsoft Corporation: Microsoft azure: cloud computing platform and services. https://azure.microsoft.com (2015). Accessed 18 Oct 2015
39. Papadopoulos, M., Mavromoustakis, C.X., Skourletopoulos, G., Mastorakis, G., Pallis, E.: Performance analysis of reactive routing protocols in mobile ad hoc networks. Proceedings of the 2014 IEEE 6th International Conference on Telecommunications and Multimedia (TEMU 2014), pp. 104–110. Heraklion, Crete, Greece (2014)
40. Park, K., Nguyen, M.C., Won, H.: Web-based collaborative big data analytics on big data as a service platform. In: Proceedings of the 2015 17th International Conference on Advanced Communication Technology (ICACT), Seoul, pp. 564–567 (2015)
41. Pop, C., Ciobanu, R., Marin, R.C., Dobre, C., Mavromoustakis, C.X., Mastorakis, G., Rodrigues, J.J.P.C.: Data dissemination in vehicular networks using context spaces. In: IEEE GLOBECOM 2015, Fourth International Workshop on Cloud Computing Systems, Networks, and Applications (CCSNA), 6–10 Dec 2015
42. Posnakides, D., Mavromoustakis, C.X., Skourletopoulos, G., Mastorakis, G., Pallis, E., Batalla, J.M.: Performance analysis of a rate-adaptive bandwidth allocation scheme in 5G mobile networks. Proceedings of the 20th IEEE Symposium on Computers and Communications (ISCC 2015)—The 2nd IEEE International Workshop on A 5G Wireless Odyssey:2020, pp. 955–961. Larnaca, Cyprus (2015)
43. Skourletopoulos, G., Xanthoudakis, A.: Developing a business plan for new technologies: application and implementation opportunities of the interactive digital (iDTV) and internet protocol (IPTV) television as an advertising tool. Bachelor's Degree Dissertation, Technological Educational Institute of Crete, Greece (2012)

44. Skourletopoulos, G.: Researching and quantifying the technical debt in cloud software engineering. Master's Degree Dissertation, University of Birmingham, UK (2013)
45. Skourletopoulos, G., Bahsoon, R., Mavromoustakis, C.X., Mastorakis, G., Pallis, E.: Predicting and quantifying the technical debt in cloud software engineering. Proceedings of the 2014 IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD 2014), pp. 36–40. Greece, Athens (2014)
46. Skourletopoulos, G., Bahsoon, R., Mavromoustakis, C.X., Mastorakis, G.: The technical debt in cloud software engineering: a prediction-based and quantification approach. In: Mastorakis, G., Mavromoustakis, C.X., Pallis, E. (eds.) Resource Management of Mobile Cloud Computing Networks and Environments, 1st edn, pp. 24–42. IGI Global, Hershey, PA (2015)
47. Skourletopoulos, G., Mavromoustakis, C.X., Mastorakis, G., Rodrigues, J.J.P.C., Chatzimisios, P., Batalla, J.M.: A fluctuation-based modelling approach to quantification of the technical debt on mobile cloud-based service level. In: Proceedings of the IEEE Global Communications Conference (GLOBECOM 2015)—The Fourth IEEE International Workshop on Cloud Computing Systems, Networks, and Applications (CCSNA), San Diego, California, USA (2015b)
48. Skourletopoulos, G., Mavromoustakis, C.X., Mastorakis, G., Pallis, E., Chatzimisios, P., Batalla, J.M.: Towards the evaluation of a big data-as-a-service model: a decision theoretic approach. In: Proceedings of the IEEE International Conference on Computer Communications (INFOCOM 2016)—First IEEE International Workshop on Big Data Sciences, Technologies and Applications (BDSTA), San Francisco, California, USA (2016a)
49. Skourletopoulos, G., Mavromoustakis, C.X., Mastorakis, G., Batalla, J.M., Sahalos, J.N.: An evaluation of cloud-based mobile services with limited capacity: a linear approach. Soft Comput. (2016). doi:10.1007/s00500-016-2083-4
50. Skourletopoulos, G., Mavromoustakis, C.X., Mastorakis, G., Pallis, E., Batalla, J.M., Kormentzas, G.: Quantifying and evaluating the technical debt on mobile cloud-based service level. In: Proceedings of the IEEE International Conference on Communications (ICC 2016)—Communication QoS, Reliability and Modeling (CQRM) Symposium, Kuala Lumpur, Malaysia (2016c)
51. Sun, X., Gao, B., Fan, L., An, W.: A cost-effective approach to delivering analytics as a service. Proceedings of the 2012 IEEE 19th International Conference on Web Services (ICWS), pp. 512–519. Honolulu, HI (2012)
52. Talia, D.: Clouds for scalable big data analytics. IEEE Comput. Sci. 98–101 (2013)
53. The Apache Software Foundation: Apache hadoop. http://hadoop.apache.org (2014). Accessed 18 Oct 2015
54. Vakintis, I., Panagiotakis, S., Mastorakis, G., Mavromoustakis, C.X.: Evaluation of a Web Crowd-sensing IoT ecosystem providing big data analysis. In: Pop, F., Kołodziej, J., di Martino, B. (eds.) Resource Management for Big Data Platforms and Applications. Studies in Big Data Springer International Publishing, 2017
55. Vaquero, L.M., Celorio, A., Cuadrado, F., Cuevas, R.: Deploying large-scale datasets on-demand in the cloud: treats and tricks on data distribution. IEEE Trans. Cloud Comput. **3**, 132–144 (2015)
56. Ye, F., Wang, Z., Zhou, F., Wang, Y., Zhou, Y.: Cloud-based big data mining and analyzing services platform integrating R. In: Proceedings of the 2013 International Conference on Advanced Cloud and Big Data (CBD), Nanjing, pp. 147–151 (2013)
57. Yeo, C.S., Venugopal, S., Chu, X., Buyya, R.: Autonomic metered pricing for a utility computing service. Future Gener. Comput. Syst. **26**, 1368–1380 (2010)
58. Zhao, Y., Calheiros, R.N., Gange, G., Ramamohanarao, K., Buyya, R.: SLA-based resource scheduling for big data analytics as a service in cloud computing environments. In: Proceedings of the 2015 44th International Conference on Parallel Processing (ICPP), Beijing, pp. 510–519 (2015)

59. Zheng, Z., Zhu, J., Lyu, M.R.: Service-generated big data and big data-as-a-service: an overview. Proceedings of the 2013 IEEE International Congress on Big Data (BigData Congress), pp. 403–410. Santa Clara, California (2013)
60. Zulkernine, F., Martin, P., Zou, Y., Bauer, M., Gwadry-Shridhar, F., Aboulnaga, A.: Towards cloud-based analytics-as-a-service (CLAaaS) for big data analytics in the cloud. Proceedings of the 2013 IEEE International Congress on Big Data (BigData Congress), pp. 62–69. Santa Clara, California (2013)

Author Biographies

Georgios Skourletopoulos is currently a Doctoral Researcher in the Mobile Systems (MoSys) Laboratory, Department of Computer Science at the University of Nicosia, Cyprus. He also works as a Junior Business Analytics and Strategy Consultant in the Business Analytics and Strategy (BA&S) Service Line within IBM's Global Business Services (GBS) Business Unit at IBM Hellas S.A., Greece and he is a member of the IBM Big Data and Business Analytics Center of Competence in Greece. He obtained his M.Sc. in Computer Science from the University of Birmingham, UK in 2013 and his B.Sc. in Commerce and Marketing (with major in e-Commerce and Digital Marketing) from the Technological Educational Institute of Crete, Greece in 2012 (including a semester as an Erasmus exchange student at the Czech University of Life Sciences Prague, Czech Republic). In the past, he has worked as an e-Banking Platforms Use Case and Quality Assurance Analyst at Scientia Consulting S.A., Greece and as a Junior Research Analyst at Hellastat S.A., Greece. Mr. Georgios Skourletopoulos has publications at various international journals, conference proceedings, workshops and book chapters. His research interests lie in the general areas of mobile cloud computing, cloud computing, cloud-based software engineering, communications and networks.

Constandinos X. Mavromoustakis is currently an Associate Professor in the Department of Computer Science at the University of Nicosia, Cyprus. He received a five-year dipl.Eng in Electronic and Computer Engineering from Technical University of Crete, Greece, his M.Sc. in Telecommunications from University College of London, UK and his Ph.D. from the Department of Informatics at Aristotle University of Thessaloniki, Greece. He serves as the Chair of C16 Computer Society chapter of the Cyprus IEEE section, whereas he is the main recipient of various grants including the ESR-EU. His research interests are in the areas of spatial and temporal scheduling, energy-aware self-scheduling and adaptive behaviour in wireless and multimedia systems.

George Mastorakis received his B.Eng. in Electronic Engineering from UMIST, UK in 2000, his M.Sc. in Telecommunications from UCL, UK in 2001 and his Ph.D. in Telecommunications from the University of the Aegean, Greece in 2008. He is serving as an Assistant Professor at the Technological Educational Institute of Crete and as a Research Associate in Research & Development of Telecommunications Systems Laboratory at the Centre for Technological Research of Crete, Greece. His research interests include cognitive radio networks, networking traffic analysis, radio resource management and energy efficient networks. He has more than 80 publications at various international conferences proceedings, workshops, scientific journals and book chapters.

Jordi Mongay Batalla received his M.Sc. degree from Universitat Politecnica de Valencia in 2000 and his Ph.D. degree from Warsaw University of Technology in 2009, where he still works as Assistant Professor. In the past, he has worked in Telcordia Poland (Ericsson R&D Co.) and later in the National Institute of Telecommunications, Warsaw, where he is the Head of Internet

Architectures and Applications Department from 2010. He took part (coordination and/or participation) in more than 10 national and international ICT research projects, four of them inside the EU ICT Framework Programmes. His research interests focus mainly on Quality of Service (Diffserv, NGN) in both IPv4 and IPv6 infrastructures, Future Internet architectures (Content Aware Networks, Information Centric Networks) as well as applications for Future Internet (Internet of Things, Smart Cities, IPTV). He is author or co-author of more than 100 papers published in books, international and national journals and conference proceedings.

Ciprian Dobre completed his Ph.D. at the Computer Science Department, University Politehnica of Bucharest, Romania, where he is currently working as a full-time Associate Professor. His main research interests are in the areas of modeling and simulation, monitoring and control of large scale distributed systems, vehicular ad hoc networks, context-aware mobile wireless applications and pervasive services. He has participated as a team member in more than 10 national projects the last four years and he was member of the project teams for 5 international projects. He is currently involved in various organizing committees or committees for scientific conferences. He has developed MONARC 2, a simulator for LSDS used to evaluate the computational models of the LHC experiments at CERN and other complex experiments. He collaborated with Caltech in developing MonALISA, a monitoring framework used in production in more than 300 Grids and network infrastructures around the world. He is the developer of LISA, a lightweight monitoring framework that is used for the controlling part in projects like EVO or the establishment of world-wide records for data transferring at SuperComputing Bandwidth Challenge events (from 2006 to 2010). His research activities were awarded with the Innovations in Networking Award for Experimental Applications in 2008 by the Corporation for Education Network Initiatives (CENIC) and a Ph.D. scholarship by Oracle between 2006 and 2008.

Spyros Panagiotakis is an Assistant Professor at the Department of Informatics Engineering of the Technological Educational Institute of Crete and leader of the Laboratory for Sensor Networks, Telematics and Industrial Information Systems. He received his Ph.D. in Communication Networks from the Department of Informatics and Telecommunications of the University of Athens, Greece in 2007. He also received a four-year fellowship for postgraduate studies from the National Center for Scientific Research "Demokritos". He has participated in several European projects (IST projects MOBIVAS, POLOS, ANWIRE, LIAISON), as well as in several national projects. He is the author of over than 40 publications in international refereed books, journals and conferences. He also serves as peer-reviewer in international conferences and journals, as well as member of technical programme committees for international conferences and workshops. Finally, he serves as evaluator of research projects for the "Competitiveness, Entrepreneurship and Innovation" Operational Programme of Greece. His research interests include mobile multimedia technologies, communications and networking, Internet of Things, pervasive computing, sensor networks, web engineering, mobile applications, automation systems, location and context awareness and informatics in education.

Evangelos Pallis is an Associate Professor in the Applied Informatics and Multimedia Department, School of Applied Technology, Technological Educational Institute of Crete, and co-director of the "PASIPHAE" Research and Development of Telecommunication Systems Laboratory. He received his B.Sc. in Electronic Engineering from the Technological Educational Institute of Crete in 1994 and his M.Sc. in Telecommunications from University of East London in 1997. He received his Ph.D. in Telecommunications from the University of East London in 2002. His research interests are in the fields of wireless networks, mobile communication systems, digital broadcasting technologies and interactive television systems, cognitive radio and dynamic access technologies, QoS/QoE techniques and network management techniques, as well as in radio-resource trading and optimisation techniques. He has participated in a number of national and European funded R&D projects, including the AC215 "CRABS", IST-2000-26298 "MAMBO", IST-2000-28521 "SOQUET", IST-2001-34692 "REPOSIT", IST-2002-FP6-507637

“ENTHRONE”, “IMOSAN”, and as technical/scientific coordinator for the IST-2002-FP6-507312 “ATHENA” project. Currently, he is involved in the FP7-214751 “ADAMANTIUM”, FP7-ICT-224287 “VITAL++” and FP7-ICT-248652 “ALICANTE” projects. He has more than 12 publications in international referred journals, 58 conference papers and 9 book chapters in the above scientific areas. He is the general chairman of the international conference on Telecommunications and Multimedia (TEMU), member of IET/IEE and active contributor to the IETF interconnection of content distribution networks (CDNi).

Springer

http://www.springer.com/978-3-319-45143-5

Advances in Mobile Cloud Computing and Big Data in the 5G Era
Mavromoustakis, C.; Mastorakis, G.; Dobre, C. (Eds.)
2017, VIII, 382 p. 149 illus., 127 illus. in color., Hardcover
ISBN: 978-3-319-45143-5